

STIC-Biotech/ChemLib

From: Slobodyansky, Elizabeth
Sent: Thursday, June 27, 2002 11:35 AM
T : STIC-Biotech/ChemLib
Subject: 09/801,852

Please search for case 09/801,852:



SEQ ID NOs: 1 and 3 against commercial ~~and interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
CM1 10D11
703-306-3222

mail box 10C01

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 6/28
Date Completed: 7/1
Searcher Prep/Review: 10
Clerical: _____
Online time: 10

TYPE OF SEARCH:
NA Sequences: 2
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 15:06:19 ; Search time 135.69 Seconds
(without alignments)
4018.766 Million cell updates/sec

Title: US-09-801-852A-3

Perfect score: 2220
Sequence: 1 ggatccaccgaactgtc.....agacgttggaagtgcgaaa 2220

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2.6/prodata/2/lna/5A-COMB.seq.*
2: /cgn2.6/prodata/2/lna/5B-COMB.seq.*
3: /cgn2.6/prodata/2/lna/6A-COMB.seq.*
4: /cgn2.6/prodata/2/lna/6B-COMB.seq.*
5: /cgn2.6/prodata/2/lna/PCTUS-COMB.seq.*
6: /cgn2.6/prodata/2/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 2220 | 100.0 | 2220 | 2 | US-08-932-376A-3 |
| 2 | 1375 | 61.9 | 1738 | 2 | US-08-932-376A-1 |
| 3 | 64.4 | 2.9 | 1635 | 2 | US-08-776-210-5 |
| 4 | 64.2 | 2.9 | 7218 | 1 | US-08-232-463-14 |
| 5 | 62.8 | 2.8 | 1828 | 3 | US-08-362-525-11 |
| 6 | 53.4 | 2.4 | 1650 | 2 | US-08-776-210-6 |
| 7 | 53.2 | 2.4 | 2539 | 3 | US-09-000-016-3 |
| 8 | 53.2 | 2.4 | 2809 | 3 | US-09-000-016-1 |
| 9 | 51 | 2.3 | 1505 | 1 | US-07-915-246-1 |
| 10 | 49.2 | 2.2 | 20235 | 3 | US-07-642-734C-3 |
| 11 | 49.2 | 2.2 | 20235 | 3 | US-08-439-009A-3 |
| 12 | 49 | 2.2 | 2064 | 1 | US-08-343-428-1 |
| 13 | 48.4 | 2.2 | 6085 | 4 | US-09-029-603-4 |
| 14 | 48.2 | 2.2 | 1931 | 2 | US-09-130-114-2 |
| 15 | 47.6 | 2.1 | 1965 | 4 | US-09-178-252-26 |
| 16 | 47.4 | 2.1 | 3196 | 4 | US-09-704-449-1 |
| 17 | 47 | 2.1 | 1485 | 4 | US-09-372-422A-39 |
| 18 | 47 | 2.1 | 44377 | 2 | US-08-804-227C-7 |
| 19 | 47 | 2.1 | 44377 | 2 | US-08-804-198-1 |
| 20 | 46.4 | 2.1 | 320 | 4 | US-09-165-264-7 |
| 21 | 45.8 | 2.1 | 1142 | 3 | US-08-765-743-1 |
| 22 | 45 | 2.0 | 3300 | 2 | US-08-194-290-6 |
| 23 | 45 | 2.0 | 3300 | 2 | US-08-614-377A-6 |
| 24 | 45 | 2.0 | 3300 | 4 | US-09-142-648B-6 |
| 25 | 45 | 2.0 | 8438 | 1 | US-07-945-283-1 |
| 26 | 44.8 | 2.0 | 1288 | 1 | US-08-440-856A-9 |
| 27 | 44.6 | 2.0 | 320 | 4 | US-09-165-264-11 |

| | | | | | | |
|------|------|-----|-------|---|-------------------|-------------------|
| C 28 | 44.6 | 2.0 | 13987 | 2 | US-08-804-227C-13 | Sequence 13, Appl |
| C 29 | 44.4 | 2.0 | 320 | 4 | US-09-165-264-14 | Sequence 14, Appl |
| C 30 | 44.4 | 2.0 | 2563 | 1 | US-08-608-241-5 | Sequence 5, Appl1 |
| C 31 | 44.4 | 2.0 | 2563 | 2 | US-08-922-182-5 | Sequence 5, Appl1 |
| C 32 | 44.4 | 2.0 | 2563 | 2 | US-08-919-953-5 | Sequence 5, Appl1 |
| C 33 | 44.2 | 2.0 | 318 | 4 | US-09-165-264-12 | Sequence 12, Appl |
| C 34 | 44.2 | 2.0 | 320 | 4 | US-09-165-264-13 | Sequence 13, Appl |
| C 35 | 44.2 | 2.0 | 1908 | 1 | US-08-173-508-1 | Sequence 1, Appl1 |
| C 36 | 44.2 | 2.0 | 1908 | 2 | US-08-265-310-1 | Sequence 1, Appl1 |
| C 37 | 44.2 | 2.0 | 1908 | 3 | US-08-951-743-1 | Sequence 1, Appl1 |
| C 38 | 43.6 | 2.0 | 1500 | 4 | US-09-593-711A-10 | Sequence 10, Appl |
| C 39 | 43.6 | 2.0 | 2418 | 1 | US-08-462-484-1 | Sequence 1, Appl1 |
| C 40 | 43.6 | 2.0 | 2418 | 1 | US-08-441-147-1 | Sequence 1, Appl1 |
| C 41 | 43.6 | 2.0 | 2418 | 5 | PCT-US95-07536-1 | Sequence 1, Appl1 |
| C 42 | 43.6 | 2.0 | 3624 | 1 | US-07-951-715A-6 | Sequence 6, Appl1 |
| C 43 | 43.6 | 2.0 | 3624 | 2 | US-08-459-448A-6 | Sequence 6, Appl1 |
| C 44 | 43.6 | 2.0 | 3624 | 3 | US-08-459-595A-6 | Sequence 6, Appl1 |
| C 45 | 43.6 | 2.0 | 3624 | 3 | US-08-459-504B-6 | Sequence 6, Appl1 |

ALIGNMENTS

RESULT 1
US-08-932-376A-3
; Sequence 3, Application US/08932376A
; Patent No. 5869309
; GENERAL INFORMATION:
; APPLICANT: Politino, Michael
; APPLICANT: Tonzi, Sean M.
; APPLICANT: Usher, John J.
; APPLICANT: Burnett K, William V.
; APPLICANT: Romancik, Guna
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: Rt. 206 & Provinceline Road
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,376A
; FILING DATE: 17-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON0144a
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4956
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-932-376A-3

Query Match 100.0%; Score 2220; DB 2; Length 2220;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of module 6"
NAME/KEY: misc_feature
LOCATION: 6054..7036
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 4"
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: dehydratase and enoylreductase domains m"
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
FEATURE: OTHER INFORMATION: beta-ketoreductase of module 4"
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyl carrier domain of module 4"
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene-eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
FEATURE: OTHER INFORMATION: 6-deoxyerythronolide B formatio"
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: module 5"
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 5"
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoreductase of module 5"
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyl carrier domain of module 5"
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: module 6"
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyltransferase domain of module 6"
NAME/KEY: misc_feature
LOCATION: 18379..18921
```

```
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: beta-ketoreductase domain of module 6"
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: acyl carrier domain of module 6"
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
FEATURE: OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match 2.2%; Score 49.2; DB 1; Length 20235;
Best Local Similarity 45.4%; Pred. No. 0.032;
Matches 216; Conservative 0; Mismatches 258; Indels 2; Gaps 1;

QY 1540 catcacctcgaggagcgccgctcggaagcagtagccgacgtcccgacgcgcgtc 1599
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17256 CACGACCGAGGAGGGGAGTTGACCGCGGACGGAGATCGGTCTCCACGGTGGAT 17197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1600 aaagggcaacacctctctgcgcatctcgccgtcgcgcgactcgcgactctgtgctg 1659
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17196 CAGCGCGGGGCGGTTTCACGGGAGCGGCAACGAGACCATGCCGCCCTTGC 17137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1660 tccccg--tgctcttcgagattatccgtgacttcccgcttgcgcgagctcccgac 1717
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17136 CGCAGCGAACGCGCTTCGCGCGCAGGGCCACGACGCGCACCGCGTCTCC 17077
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1718 ctactgagcgcgagcggttcggtctgcgcgcacaaaggccctcttcgactacgcgc 1777
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17076 CCGCGCCACACCGCGCGCGGATCTGCCCTCGGATGCCGATGACGCGCGG 17017
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1778 ggtcaccagcgaccgacaaactcgtactacatcgggtccatcttggaacggaaagtc 1837
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17016 GACTCCGAAAGCCGCCACAGCGCGGCGACGACACCATCAGGAGAACACAC 16957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1838 ggtctcgtcgtccagtccttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16956 TAGACGCTGACCGCTCCAGCGACGCGCGGCTCGGACGCTGCTCCAGAC 16897
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1898 gaacaacaagcgtgccacaagaccatcaacccttactggccgacgttcgactcgggcaa 1957
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16896 CGAGAACCGCGCCACCTCCGACACACCGCATCGCATCGCGGATCGACTCG 16837
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1958 gcagctcctcttaacacgacgacgagggacacctctctcccgccgacccgcga 2013
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16836 GGGGACCGAGAGAAAGCCCGGGCCATGCCCTCCACTGAGCACCCCTGAC 16781
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-439-009A-3/c
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18375..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
```



```
; SEQ ID NO 26
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic B.t. toxin gene
US-09-178-252-26
```

```
Query Match      2.18; Score 47.6; DB 4; Length 1965;
Best Local Similarity 46.78; Pred. No. 0.036;
Matches 227; Conservative 0; Mismatches 249; Indels 10; Gaps 2;

Qy 1493 cagcggtctccagcttcgagcgcctctccgagcctctctccctacatcacctcggag 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 ccgcgcgccaaccgggcatctgcgcgcgtggggaacaaccggtacaacacctctcc 1011

Qy 1553 gagcgccagcgctgcgaagcagctacccgagctctccgagcgcgcgctcaagggaacacc 1612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1012 gagctggagaagccttcacagcgccgcgcacctcttcgagcgctcaacgcgcctacc 1071

Qy 1613 ttctctgcacatcgccgctacgcgcgactcgacctctgctgctgcttcccgctgctt 1672
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1072 atctccgcgaaccgtacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1124

Qy 1673 ctccgagtgattccgctgacttccgcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1125 gtcgggccacacctgcagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1184

Qy 1733 ggcgttcggctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1185 cggccagatcacctcca---acacccgcctcttcacacccacacacgcgcgcgcgcgcgc 1241

Qy 1793 cgacaactcgtactacatcgctccatctggaacgcgcaagaagtcgctcgcgcgcgcgc 1852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1242 cgactccaggcgcgcaatttcggcaacctctacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1301

Qy 1853 gtccttcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1302 caacatcttcgcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1361

Qy 1913 caacaagaccatcaacccttactgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1362 ggacctcaccacccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1421

Qy 1973 caccgac 1978
Db 1422 cgtgac 1427
```

Search completed: June 29, 2002, 15:07:14
Job time: 20670 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 13:05:40 ; Search time 4149.72 seconds
(without alignments)
7220.545 Million cell updates/sec

Title: US-09-801-852A-3
Perfect score: 2220
Sequence: 1 ggatccaccggaactctgtc.....agacgttgcaagtgcgaaa 2220

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estma:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_esti:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|------------|---------------------|
| C 1 | 64.2 | 2.9 | 828 | CNS04GHZ | AL289664 Tetraodon |
| C 2 | 60.6 | 2.7 | 793 | CNS01T34 | AL166009 Tetraodon |
| C 3 | 59.8 | 2.7 | 1101 | CNS016MU | AL106944 Drosophila |
| C 4 | 59.4 | 2.7 | 1101 | CNS00GZV | AL072957 Drosophila |
| C 5 | 59.2 | 2.7 | 625 | AA202160 | AA202160 LD02542.5 |
| C 6 | 59.2 | 2.7 | 649 | AA246278 | AA246278 LD05065.5 |
| C 7 | 59.2 | 2.7 | 673 | B1574042 | B1574042 RH09457.5 |
| C 8 | 59.2 | 2.7 | 676 | B1564550 | B1564550 RH61852.5 |
| C 9 | 58.6 | 2.6 | 880 | CNS04463 | AL273684 Tetraodon |
| C 10 | 58.4 | 2.6 | 925 | CNS0091P | AL053013 Drosophila |
| C 11 | 58.2 | 2.6 | 566 | CNS03JNO | AL247077 Tetraodon |
| C 12 | 58.2 | 2.6 | 1463 | BM320858 | BM320858 rumast-1 |
| C 13 | 58 | 2.6 | 1464 | 3 BE636746 | BE636746 rockefell |
| C 14 | 57.6 | 2.6 | 1011 | AG126307 | AG126307 Pan trogl |
| C 15 | 57.2 | 2.6 | 568 | AI516721 | AI516721 LD42723.5 |
| C 16 | 57.2 | 2.6 | 855 | AL572700 | AL572700 AL572700 |
| C 17 | 57.2 | 2.6 | 873 | AG136937 | AG136937 Pan trogl |

| | | | | |
|------|------|-----|------|-------------|
| C 18 | 57 | 2.6 | 631 | 10 BI955992 |
| C 19 | 57 | 2.6 | 917 | 12 AG081214 |
| C 20 | 56.8 | 2.6 | 857 | 12 AG044313 |
| C 21 | 56.4 | 2.5 | 586 | 10 BI626691 |
| C 22 | 56.4 | 2.5 | 703 | 12 AQ738630 |
| C 23 | 56.4 | 2.5 | 1101 | 12 CNS0153F |
| C 24 | 56.2 | 2.5 | 829 | 9 AW448067 |
| C 25 | 56.2 | 2.5 | 988 | 12 AG135153 |
| C 26 | 56.2 | 2.5 | 1321 | 12 AG126084 |
| C 27 | 56 | 2.5 | 1037 | 12 AG146864 |
| C 28 | 56 | 2.5 | 1101 | 12 CNS001T4 |
| C 29 | 55.8 | 2.5 | 781 | 12 AG153560 |
| C 30 | 55.6 | 2.5 | 1223 | 10 BI416573 |
| C 31 | 55.4 | 2.5 | 867 | 10 BI958202 |
| C 32 | 55.4 | 2.5 | 1064 | 12 AG134261 |
| C 33 | 55.4 | 2.5 | 1083 | 10 BM460150 |
| C 34 | 55.4 | 2.5 | 1101 | 12 CNS01523 |
| C 35 | 55.2 | 2.5 | 768 | 12 AG138112 |
| C 36 | 55.2 | 2.5 | 932 | 12 AG134650 |
| C 37 | 55.2 | 2.5 | 1052 | 10 BI416480 |
| C 38 | 55.2 | 2.5 | 1082 | 10 BE455154 |
| C 39 | 55.2 | 2.5 | 1155 | 12 AG124478 |
| C 40 | 55 | 2.5 | 820 | 10 BG320478 |
| C 41 | 55 | 2.5 | 976 | 10 BI416536 |
| C 42 | 54.8 | 2.5 | 809 | 10 BI952732 |
| C 43 | 54.8 | 2.5 | 878 | 12 AZ183970 |
| C 44 | 54.8 | 2.5 | 926 | 12 CNS0364D |
| C 45 | 54.8 | 2.5 | 933 | 12 AG126371 |

ALIGNMENTS

RESULT 1
CNS04GHZ/c
LOCUS
DEFINITION
CNS04GHZ 828 bp DNA linear GSS 21-MAY-2000
Tetraodon nigroviridis genome survey sequence pUC-ori end of clone
108D17 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL289664
VERSION
AL289664.1 GI:8028243
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 828)
Roest-Crollius H., Jaillon O., Dasilva C., Fizames C., Fisher C.,
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and
Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL
2 (bases 1 to 828)
Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL
3 (bases 1 to 828)
Genoscope.
Direct Submission
TITLE
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
Location/Qualifiers
source
1..828
/organism="Tetraodon nigroviridis"


```
Query Match      2.7%; Score 59.2; DB 9; Length 625;
Best Local Similarity 59.2%; Pred. No. 1.2;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgcgaggcgacaat 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 305 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTACGCGGACTCCT 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 494 cttccgctcctgtctacattcacaggagtgctcagccttcggcgatcgagcaccggc 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 365 TTACCGGTGATGCTGTGGATTTCATGTGGAGGCTTCTTCTCGCAACGCAACAGTGAC 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 554 agcgacttcgcttcaccacgacagcggaacaaagatggtcgtgttaaatctccag 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 425 TTCCACTTTCCCG-----CCAAGCTCATGAGGAGGAGTTCATGTGTGTCACCTGTAAT 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 614 taccgtctcggcagctttggtttcct 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 479 TACCGACTTGGAGCGCTTGGTTTCT 504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
LOCUS      AA246278      649 bp      mRNA      linear      EST 19-APR-2001
DEFINITION Drosophila melanogaster embryo Bluescript
ACCESSION  AA246278
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 649)
AUTHORS    Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 50 row: F column: 5
            High quality sequence stop: 539.
            Location/Qualifiers
                1..649
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="LD05065"
                /clone.lib="LD Drosophila melanogaster embryo Bluescript"
                /sex="male and female"
                /dev_stage="0 to 24 hours mixed stage embryonic"
                /lab_host="SOLR"
                /note="Organ: embryo; Vector: Bluescript SK; Site:1: EcoRI
                ; Site:2: XhoI; Constructed using Stratagene ZAP-cDNA
                Synthesis kit. Oligo dt-primed and directionally cloned at
                EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT  153 a 150 c 184 g 161 t 1 others
ORIGIN
```

```
Query Match      2.7%; Score 59.2; DB 9; Length 649;
Best Local Similarity 59.2%; Pred. No. 1.2;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgcgaggcgacaat 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 351 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTACGCGGACTCCT 410
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 494 cttccgctcctgtctacattcacaggagtggtcagccttcggcgatcgagcaccggc 553
```

```
Db 411 TTACCGGTGATGCTGTGGATTTCATGTGGAGGCTTCTTCTCGCAACGCAACAGTGAC 470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 554 agcgacttcgcttcaccacgacagcggaacaaagatggtcgtgttaaatctccag 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 471 TTCCACTTTCCCG-----CCAAGCTCATGAGGAGGAGTTCATGTGTGTCACCTGTAAT 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 614 taccgtctcggcagctttggtttcct 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 525 TACCGACTTGGAGCGCTTGGTTTCT 550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
LOCUS      BI574042      673 bp      mRNA      linear      EST 06-SEP-2001
DEFINITION RH09457 SpRime RH Drosophila melanogaster normalized Head pf1c-1
            Drosophila melanogaster cDNA clone RH09457 5 similar to c1t:
            FBan000958 GO: [carboxyesterase (GO:0004091); carboxyesterase
            (GO:0004091)] located on: 2R 57F4-57F4; : 07/26/2001, mRNA sequence.
ACCESSION  BI574042
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 673)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,B., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
TITLE      BDGP/HMI RH Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Other_ESTs: RH09457.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003454: arm:2R [16312727,16613591]
            estimated-cyto:57D11-57F8: 07/26/2001
            Plate: RH.94 row: E column: 9
            High quality sequence stop: 577.
            Location/Qualifiers
                1..673
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="RH09457"
                /clone.lib="RH Drosophila melanogaster normalized Head
                pf1c-1"
                /sex="male and female"
                /dev_stage="Adult"
                /lab_host="DH5-alpha Tona"
                /note="Organ: head; Vector: pf1c1; Site:1: XhoI; Site:2:
                BamHI; Library was kindly generated by Piero Carninci at
                the RIKEN. The library was normalized and excised using
                Cre recombinase. Plasmid cDNA library."
BASE COUNT  157 a 156 c 191 g 168 t 1 others
ORIGIN
```

```
Query Match      2.7%; Score 59.2; DB 10; Length 673;
Best Local Similarity 59.2%; Pred. No. 1.2;
Matches 122; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 434 tcggaggattgctcttctcaatgctgtgccccgcggctctgtgcgaggcgacaat 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 374 TCCGAGGACTGCTTTCTTCTTATGTATATGACCTAAAGTACGCGGACTCCT 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

| | | | |
|----|-----|---|-----|
| Qy | 494 | cttccgcgtctcgtctacattcaacgagaggtggtacgcgtcttcgagcatgcgagcaacgcgc | 553 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 434 | TTACCCGTGATGGTCTGGATTATGTTGGAGGCTTCTTCTCGCAACGGCAACAGTGAC | 493 |
| | | | |
| | | | |
| | | | |
| | | | |
| Qy | 554 | agcgactttgccgccttccaccacacacacgcgggaacaaagatggttcgttgttaaatctccag | 613 |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 494 | TTCCACTTTCCCG-----CCAAGCTCATGGACGAGGAGTTCATTGTGTCAACCCCTGAAT | 547 |
| | | | |
| | | | |
| | | | |
| | | | |
| Qy | 614 | taccgctctcggagcgtttggtttctct | 639 |
| | | | |
| | | | |
| | | | |
| Db | 548 | TACCGACTTTGGAGCGCTTTGTTTCTCT | 573 |

| | | |
|------------|---|------------------|
| RESULT | 9 | |
| CNS04463/c | | |
| LOCUS | 880 bp | DNA |
| DEFINITION | Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone 080p10 of library G from Tetraodon nigroviridis, genomic survey sequence. | |
| ACCESSION | AL273684 | |
| VERSION | AL273684.1 | GI:7995946 |
| KEYWORDS | GSS; genome survey sequence. | |
| SOURCE | Tetraodon nigroviridis. | |
| ORGANISM | Tetraodon nigroviridis | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. | |
| REFERENCE | 1 | (bases 1 to 880) |
| AUTHORS | Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. | |
| TITLE | Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis | |
| JOURNAL | Unpublished | |
| REFERENCE | 2 | (bases 1 to 880) |
| AUTHORS | Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,C., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. | |

| | | | | |
|------------|--|------------------|-------|-----------|
| JOURNAL | unpublished Tetraodon nigroviridis DNA sequence | | | |
| REFERENCE | 3 | (bases 1 to 880) | | |
| AUTHORS | Genoscope. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases | | | |
| COMMENT | This sequence is a single read and was generated as part of a scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon . | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..880 | | | |
| | /organism="Tetraodon nigroviridis" | | | |
| | /db_xref="taxon:99893" | | | |
| | /clone="080P10" | | | |
| | /clone_lib="C" | | | |
| | /note="Genoscope sequence ID : COBG080DH05SP1-end : PUC-ori" | | | |
| BASE COUNT | 87 a | 96 c | 518 g | 109 t |
| ORIGIN | | | | 70 others |

| | 137, 138 | 139, 140 | 141, 142 | 143, 144 | 145, 146 | 147, 148 | 149, 150 | 151, 152 | 153, 154 | 155, 156 | 157, 158 | 159, 160 | 161, 162 | 163, 164 | 165, 166 | 167, 168 | 169, 170 | 171, 172 | 173, 174 | 175, 176 | 177, 178 | 179, 180 | 181, 182 | 183, 184 | 185, 186 | 187, 188 | 189, 190 | 191, 192 | 193, 194 | 195, 196 | 197, 198 | 199, 200 | 201, 202 | 203, 204 | 205, 206 | 207, 208 | 209, 210 | 211, 212 | 213, 214 | 215, 216 | 217, 218 | 219, 220 | 221, 222 | 223, 224 | 225, 226 | 227, 228 | 229, 230 | 231, 232 | 233, 234 | 235, 236 | 237, 238 | 239, 240 | 241, 242 | 243, 244 | 245, 246 | 247, 248 | 249, 250 | 251, 252 | 253, 254 | 255, 256 | 257, 258 | 259, 260 | 261, 262 | 263, 264 | 265, 266 | 267, 268 | 269, 270 | 271, 272 | 273, 274 | 275, 276 | 277, 278 | 279, 280 | 281, 282 | 283, 284 | 285, 286 | 287, 288 | 289, 290 | 291, 292 | 293, 294 | 295, 296 | 297, 298 | 299, 300 | 301, 302 | 303, 304 | 305, 306 | 307, 308 | 309, 310 | 311, 312 | 313, 314 | 315, 316 | 317, 318 | 319, 320 | 321, 322 | 323, 324 | 325, 326 | 327, 328 | 329, 330 | 331, 332 | 333, 334 | 335, 336 | 337, 338 | 339, 340 | 341, 342 | 343, 344 | 345, 346 | 347, 348 | 349, 350 | 351, 352 | 353, 354 | 355, 356 | 357, 358 | 359, 360 | 361, 362 | 363, 364 | 365, 366 | 367, 368 | 369, 370 | 371, 372 | 373, 374 | 375, 376 | 377, 378 | 379, 380 | 381, 382 | 383, 384 | 385, 386 | 387, 388 | 389, 390 | 391, 392 | 393, 394 | 395, 396 | 397, 398 | 399, 400 | 401, 402 | 403, 404 | 405, 406 | 407, 408 | 409, 410 | 411, 412 | 413, 414 | 415, 416 | 417, 418 | 419, 420 | 421, 422 | 423, 424 | 425, 426 | 427, 428 | 429, 430 | 431, 432 | 433, 434 | 435, 436 | 437, 438 | 439, 440 | 441, 442 | 443, 444 | 445, 446 | 447, 448 | 449, 450 | 451, 452 | 453, 454 | 455, 456 | 457, 458 | 459, 460 | 461, 462 | 463, 464 | 465, 466 | 467, 468 | 469, 470 | 471, 472 | 473, 474 | 475, 476 | 477, 478 | 479, 480 | 481, 482 | 483, 484 | 485, 486 | 487, 488 | 489, 490 | 491, 492 | 493, 494 | 495, 496 | 497, 498 | 499, 500 | 501, 502 | 503, 504 | 505, 506 | 507, 508 | 509, 510 | 511, 512 | 513, 514 | 515, 516 | 517, 518 | 519, 520 | 521, 522 | 523, 524 | 525, 526 | 527, 528 | 529, 530 | 531, 532 | 533, 534 | 535, 536 | 537, 538 | 539, 540 | 541, 542 | 543, 544 | 545, 546 | 547, 548 | 549, 550 | 551, 552 | 553, 554 | 555, 556 | 557, 558 | 559, 560 | 561, 562 | 563, 564 | 565, 566 | 567, 568 | 569, 570 | 571, 572 | 573, 574 | 575, 576 | 577, 578 | 579, 580 | 581, 582 | 583, 584 | 585, 586 | 587, 588 | 589, 590 | 591, 592 | 593, 594 | 595, 596 | 597, 598 | 599, 600 | 601, 602 | 603, 604 | 605, 606 | 607, 608 | 609, 610 | 611, 612 | 613, 614 | 615, 616 | 617, 618 | 619, 620 | 621, 622 | 623, 624 | 625, 626 | 627, 628 | 629, 630 | 631, 632 | 633, 634 | 635, 636 | 637, 638 | 639, 640 | 641, 642 | 643, 644 | 645, 646 | 647, 648 | 649, 650 | 651, 652 | 653, 654 | 655, 656 | 657, 658 | 659, 660 | 661, 662 | 663, 664 | 665, 666 | 667, 668 | 669, 670 | 671, 672 | 673, 674 | 675, 676 | 677, 678 | 679, 680 | 681, 682 | 683, 684 | 685, 686 | 687, 688 | 689, 690 | 691, 692 | 693, 694 | 695, 696 | 697, 698 | 699, 700 | 701, 702 | 703, 704 | 705, 706 | 707, 708 | 709, 710 | 711, 712 | 713, 714 | 715, 716 | 717, 718 | 719 |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----|

[illegible]

| | | |
|------------|----|--|
| RESULT | 10 | |
| CNS0091P/c | | |
| LOCUS | | |
| DEFINITION | | |
| ACCESION | | |
| VERSION | | |
| KEYWORDS | | |
| SOURCE | | |
| ORGANISM | | |
| REFERENCE | | |
| AUTHORS | | |
| TITLE | | |
| JOURNAL | | |
| COMMENT | | |

```

found at http://bio.cpa.msu.edu/urotophilia_bac.html.
FEATURES
  source
    Location/Qualifiers
      1..925
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RpCI-98"
        /clone="BACR19D16"
        /note="end : TET3"
BASE COUNT      120 a      61 c      61 g      172 t      511 others
ORIGIN
Query Match      2.6%; Score 58.4; DB 12; Length 925;
Best Local Similarity 11.5%; Pred. No. 1.7;
Matches 41; Conservative 178; Mismatches 139; Indels 0; Gaps 0;

```

[illegible]

| | | |
|-----|---|-----|
| 859 | ACVKCNASSCGCCGCGMACCWCSSSSCCGASARGVVRASGGAGKRGSGGASAS | 800 |
| 183 | ccgtcgtcgacctgggtacgcgcgcgtaccaaagctacttgaacgagacccgcgactct | 242 |
| 799 | HSSSACBSSSSSCASCWASSSSSASRRSRGGGAGSGASSRSSSSSSASAGSV | 740 |
| 243 | actgtgtgcgcgaatccgtacgccttcgctcagcgtctccaggtctctcagacgcgc | 302 |
| 739 | VSSASSSSSCSSSVSCSVASSMSCSSBSSSSASASSSSSSSSASCSCCCTSWSCS | 680 |
| 303 | cgacgcacaaggcgtccgcgaacgcactagatgtagcgcagtctgtgcgcgctagcg | 362 |
| 679 | CSYASMSAARSSSSSSSCSSMSASASSSSASASSSSSSSSSSSGACBSWSGG | 620 |
| 363 | agggaaaccaacagacaaaggttgcgcgcgcctagacaacgtcgagcgcgc | 420 |
| 619 | GSQSVASSGMSVSSSSGSRSGSGGGGGVGGSSSGSSGSGSGSGVCSGSGGCM | 562 |

| | |
|------------|--|
| RESULT | 11 |
| CNS03JN0/c | |
| LOCUS | |
| DEFINITION | CNS03JN0 566 bp DNA linear GSS 17-MAY-2000 Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone Q31B04 of library G from Tetraodon nigroviridis, genomic survey sequence. |
| ACCESSION | AL247077 |
| VERSION | AL247077.1 GI:7968089 |
| KEYWORDS | GSS; genome survey sequence. |
| SOURCE | Tetraodon nigroviridis. |
| ORGANISM | Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon. |

| | |
|-----------|---|
| REFERENCE | 1 (bases 1 to 566) |
| AUTHORS | Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. |
| TITLE | Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 566) |
| AUTHORS | Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J. |
| TITLE | Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence |
| JOURNAL | Unpublished |
| REFERENCE | 3 (bases 1 to 566) |

| | |
|---------|--|
| AUTHORS | Genoscope. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases |
| COMMENT | This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon . |

| FEATURES | source | Location/Qualifiers |
|------------|--------|--|
| | | 1. .566 |
| | | /organism="Tetraodon nigroviridis" |
| | | /db_xref="taxon:99883" |
| | | /clone="Q31B04" |
| | | /clone_lib="C" |
| | | /note="Genoscope sequence ID : C0BG031DA02SP1-end : PUC-ori" |
| BASE COUNT | 51 a | 52 c 332 g 70 t 61 others |
| ORIGIN | | |

| | | | | |
|-----------------------|--------------|---------------|----------------|------------|
| Query Match | 2.6% | Score 58.2 | DB 12 | Length 566 |
| Best Local Similarity | 42.6% | Pred. No. 1.7 | | |
| Matches 146 | Conservative | 34 | Mismatches 162 | Indels 1 |
| | | | | Gaps 1 |

| FEATURES | Location/Qualifiers |
|------------|---|
| source | 1. 568 |
| | /organism="Drosophila melanogaster" |
| | /db_xref="taxon:7227" |
| | /clone="LD42723" |
| | /clone_lib="LD Drosophila melanogaster embryo pOT2" |
| | /sex="male and female" |
| | /dev_stage="0 to 24 hours mixed stage embryonic" |
| | /lab_host="X11 Blue" |
| | /note="Organ: embryo; Vector: pOT2; Site_1: EcORI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2." |
| BASE COUNT | 129 a 131 c 159 g 149 t |

[illegible]

Search completed: June 29, 2002, 13:05:51
Job time: 13722 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 15:03:51 ; Search time 7060.77 seconds
(without alignments)
6579.581 Million cell updates/sec

Title: US-09-801-852a-3

Perfect score: 2220

Sequence: 1 ggatccaccggaactctgtc.....agacgttgcaagtcgaaa 2220

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match Length | ID | Description |
|------------|-------|--------------|----|-------------|
|------------|-------|--------------|----|-------------|

| | | | | | |
|----|--------|-------|--------|---|----------|
| 1 | 2220 | 100.0 | 2220 | 6 | AR034145 |
| 2 | 2220 | 100.0 | 2220 | 6 | AR034145 |
| 3 | 2220 | 100.0 | 2220 | 6 | BD005939 |
| 4 | 1375 | 61.9 | 1738 | 6 | AR034144 |
| 5 | 1375 | 61.9 | 1738 | 6 | BD005938 |
| 6 | 1373.4 | 61.9 | 1738 | 6 | AF025410 |
| 7 | 1353 | 60.9 | 1716 | 6 | AX268011 |
| 8 | 74.2 | 3.3 | 298166 | 2 | AC087563 |
| 9 | 73.8 | 3.3 | 165556 | 2 | AC087190 |
| 10 | 73.4 | 3.3 | 303091 | 2 | AC084799 |
| 11 | 71.8 | 3.2 | 220469 | 2 | AC074307 |
| 12 | 69.8 | 3.1 | 181988 | 2 | AC090552 |
| 13 | 68 | 3.1 | 150120 | 8 | AP001552 |
| 14 | 67.8 | 3.1 | 171574 | 2 | AC012300 |
| 15 | 67.6 | 3.0 | 1635 | 8 | GC002524 |
| 16 | 67.4 | 3.0 | 125020 | 9 | AF429315 |
| 17 | 66 | 3.0 | 1635 | 8 | GC002622 |
| 18 | 66 | 3.0 | 1692 | 6 | E02678 |
| 19 | 64.4 | 2.9 | 1635 | 6 | A48376 |
| 20 | 64.4 | 2.9 | 1635 | 8 | GC002387 |
| 21 | 64.4 | 2.9 | 1855 | 8 | CCLIP3 |
| 22 | 64.2 | 2.9 | 7218 | 6 | I66494 |
| 23 | 63.4 | 2.9 | 129585 | 8 | AC090120 |
| 24 | 62.8 | 2.8 | 1635 | 8 | GCTAGL |
| 25 | 62.8 | 2.8 | 1635 | 8 | GC002525 |
| 26 | 62.8 | 2.8 | 1635 | 8 | GC002541 |
| 27 | 62.8 | 2.8 | 1635 | 8 | GC002625 |
| 28 | 62.8 | 2.8 | 1828 | 6 | A74255 |
| 29 | 61.4 | 2.8 | 10732 | 6 | E32986 |
| 30 | 61.2 | 2.8 | 1635 | 8 | GC002623 |
| 31 | 61.2 | 2.8 | 1674 | 6 | E02497 |
| 32 | 61.2 | 2.8 | 1767 | 8 | GALIP2 |
| 33 | 61.2 | 2.8 | 1783 | 8 | AB000260 |
| 34 | 61.2 | 2.8 | 1812 | 8 | CCLIP5 |
| 35 | 59.4 | 2.7 | 265537 | 2 | AC087228 |
| 36 | 59.2 | 2.7 | 1834 | 3 | AF233526 |
| 37 | 59.2 | 2.7 | 1860 | 3 | AY051959 |
| 38 | 59.2 | 2.7 | 29705 | 2 | AC097965 |
| 39 | 59.2 | 2.7 | 150332 | 2 | AC096099 |
| 40 | 58.6 | 2.6 | 102242 | 2 | AP004043 |
| 41 | 58.4 | 2.6 | 78220 | 2 | AC023212 |
| 42 | 58.2 | 2.6 | 45313 | 1 | SCD95A |
| 43 | 58.2 | 2.6 | 135119 | 2 | AC011578 |
| 44 | 58 | 2.6 | 2043 | 8 | CCLIP2 |
| 45 | 57 | 2.6 | 154409 | 2 | AP003724 |

ALIGNMENTS

| | | | | | | |
|-----------------------|---|------------------------------------|---------|--------------|--------|-----------------|
| RESULT 1 | AR034145 | Sequence 3 from patent US 5869309. | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| LOCUS | AR034145 | Sequence 3 from patent US 5869309. | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| DEFINITION | AR034145 | Sequence 3 from patent US 5869309. | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| ACCESSION | AR034145 | Sequence 3 from patent US 5869309. | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| VERSION | AR034145.1 | GI:5949750 | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| KEYWORDS | AR034145.1 | GI:5949750 | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| SOURCE | Unknown. | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| ORGANISM | Unknown. | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| REFERENCE | 1 (bases 1 to 2220) | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| AUTHORS | Polittino, M., Tonzi, S.M., Usher, J.J., Burnett, W.V. and Romancik, G. | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| TITLE | Cephalosporin esterase gene from Rhodospiridium toruloides | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| JOURNAL | Patent: US 5869309-A 3 09-FEB-1999; | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| FEATURES | Location/Qualifiers | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| source | 1..2220 | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| BASE COUNT | 389 a 814 c 554 g 463 t | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| ORIGIN | /organism="unknown" | | 2220 bp | DNA | linear | PAT 29-SEP-1999 |
| Query Match | 100.0%; | Score 2220; | DB 6; | Length 2220; | | |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | | | |

Matches 2220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| Qy | 1 | ggaatcaaccggaactctgtctcccggtttctcggctttctcttctctgtctgtcgcccatcgccct | 60 |
| Db | 1 | GGATCACCCGGAACCTCTGTCGCGCTTTCTGGCTTTCTTCCTTGTGTGTCGCGCCCATCGCCT | 60 |
| Qy | 61 | ttcccgactcgcgcgaatcgtctcttaacctcttcaacctcgctccctcgtctcgagcgt | 120 |
| Db | 61 | TTCCCGACTCGCGGCATGCTCCTTAACCTCTTCACCCTCGCCTCGCTCTCGTCGGAGCT | 120 |
| Qy | 121 | ceagctcgctttgctctctcgaacctcctctgctcgccgcgcagcaaccccaaacgagccccc | 180 |
| Db | 121 | CCAGCTCGCCTTTTGGCTCTCTCGACCTCCCTGTCGCGCGACGAAACCAACAGAGCGCCC | 180 |
| Qy | 181 | ttccgctgcgcacctcgcgtacgcgcctacaaaggctactttaacgagaccgccgagct | 240 |
| Db | 181 | TCCGCTGCTGACACTCGGCTACGCCGCTACAAAGCTACTTTGAACGAGACCGCCGGACT | 240 |
| Qy | 241 | ctactggtggcgcggaatccgctacgcctcggtcagcgttccaggtcctctcaagcgc | 300 |
| Db | 241 | CTACTGGTGGCGGGAATCCGCTACGCTCGGCTCAGCGTTCCAGGCTCCTCAGACGCC | 300 |
| Qy | 301 | cgcgacgcacaaaggccgctccgcaacgcactgagtagtagaacgatactgttggccggtcag | 360 |
| Db | 301 | CGCGACGCACAAGCGCTCGCAACGCGACTGAGTAGTGACCGCATCTTTGGCCGGCTAG | 360 |
| Qy | 361 | cgaggaacaaacacagacaaagggttgcgcgcgcctagcaaacagctcgcgacgacgcgc | 420 |
| Db | 361 | CGAGGGAACCAACAGACCAAGGGCTTCGCGCGCCTAGCAACAGCTCGAGCAGCGGCC | 420 |
| Qy | 421 | gcagaaacaggcgtcggaggaattgcctcttcccaatgtcgttgcctcccccgcgcgtcgtg | 480 |
| Db | 421 | GCAGAAACAGCGGTCGAGAGATTGCTCTTCTCAATGCTTCCTGCCCGCCGCGCTCGTG | 480 |
| Qy | 481 | cgaggcgacaattctccgcctcctcgctctacatcacgaggttggtctaccccttcgacga | 540 |
| Db | 481 | CGAGGGGACAAATCTTCCCGTCTCTGTCTACATTACGGAGGTGGCTACGCTTTCGGCGA | 540 |
| Qy | 541 | tgcgagcacgcgcgagcactttgcgccttcaccaagcacacgcgggaaccagaatggtcgt | 600 |
| Db | 541 | TGCGAGCACCGGCAGCGACTTTGCCGCTTCACCAAGCACACGGGAACCAAGATGGTCTGT | 600 |
| Qy | 601 | tgtaaatctccagtaacgctctcggcagctttggtttctcgtcgtggcaacgcacatgaagga | 660 |
| Db | 601 | TGTAAATCTCCAGTACCGTCTCGGACGCTTTGGTTTCTCTGCTGGCCAAACCATGAAGGA | 660 |
| Qy | 661 | ctacggtgtaaacgaacgcgcgcttgcctgacacaggtgagttcccgcatgataccgcgcc | 720 |
| Db | 661 | CTACGGTGTAAACGACCGCGGCTTGTGTACCAAGGTGAGTTTCCGGCATGATACCCGCC | 720 |
| Qy | 721 | accttgcgactcatgctgcgcctctcccgctcgcagcaattcccttcaatgggttca | 780 |
| Db | 721 | ACCTTTGCACTCATGCTGACGCTCTCCGCTCGCAGCAATTCCGCCCTTCAATGGGTCA | 780 |
| Qy | 781 | acagcagctctgaattcggcgcgaaccccgatcaccttacgatttgggcgcagctcgc | 840 |
| Db | 781 | ACAGCAGCTCTCAAGTTCCGGCGGAACCCCGATCACGTTACGATTTGGGGCGAGTCTGC | 840 |
| Qy | 841 | agcgcagggtccgcttatgaaccagatcattgcgaacgtgagccacccgaaccgatccc | 900 |
| Db | 841 | AGCGCAGGGTCCGTTATGAACAGATCATTTGCGAAGCTGAGCCACCGAACCGATCTCC | 900 |
| Qy | 901 | agccgactttcc | 960 |
| Db | 901 | AGCCGACTTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCTCTTTCGAGGGCGGCAACA | 960 |
| Qy | 961 | ccgtcaaggctctcggctcgaagaagcccccttccacgctgcacatcggtcctccgctct | 1020 |
| Db | 961 | CCGTCGAAGGCTCTCGGTCTCAAGAGGCCCTCTTCCACGCTGCCATCGGCTCCTCCGCTCT | 1020 |
| Qy | 1021 | tctctccctaccgaacgaagtacaactcccccttcgcgcagctgctctactctcccaactcg | 1080 |
| Db | 1021 | TCCTCCCTACCAAGCAAGTACAATCCCCCTTCGCGAGCTGCTACTCTACTCCAACTCG | 1080 |

| | | | |
|----|------|--|------|
| Qy | 1081 | tctcgcgcaaaactgtacccaaagcgccctgttcttcgttgctctgaagctgtcgacg | 1140 |
| Db | 1081 | TCTCGGCGACAAACTGCACAAAGCGCGTCTGCTTCGTTCGGCTTGCGTGAAGCTGTCTGACG | 1140 |
| Qy | 1141 | ctcgcgcgtcactgcgcgcggcgtgaagaacctcggcgcttcccgttcgggttttsgt | 1200 |
| Db | 1141 | CTCGGCGCTCGCTGCGCGGGCGTGAAGAATCTGGCGGCCGTTCCCGTTCGGGTTTTGGT | 1200 |
| Qy | 1201 | cgtatgtccccggtcgtcgacgggaacctctttagctgaagcgcgctcgtctctctetgcc | 1260 |
| Db | 1201 | CGTAATGTCCCGETGCTGCAGCGGSACTTCTTAGCTAGACGCGCTGCTCTCCTTCGCCA | 1260 |
| Qy | 1261 | aggccaagaaacctcaatggcgtgcgtgycgaagctttagtgcttcagagatctcgt | 1320 |
| Db | 1261 | AGGCRAGAAGAACCTCAATGGCGTGCGTGGCGAGCTTTCAGATGCTTCAGGATCTCGCT | 1320 |
| Qy | 1321 | gacactgtcacccgctcgcagaaacctctaccgggatcaacaacctcgcacgaagtga | 1380 |
| Db | 1321 | GACACTGTGACCCTGCTCGAGAACCTTCTCACCGGATCAACAACCTCGACGAAGATGA | 1380 |
| Qy | 1381 | gttccgtcgacggtctgttcgcccgagagactgttcttttgcgaagattacg | 1440 |
| Db | 1381 | GTTCGCTGACGSGCTCTGTTCGCCAGGAGACTGACTTGTCTCTTTGCGAAGATTACG | 1440 |
| Qy | 1441 | atctaatctaactgagccactatttaagaacagacagatcagcgaccaagtgcgacgcgt | 1500 |
| Db | 1441 | ATTCATATTCACTAGCGCCACTATTTCAGAACACACGATCAGCGACCAGTCCGACGCGCT | 1500 |
| Qy | 1501 | ctcccagttcacgcctcctgcgcgcctcttccccatacatcaactcggagggagcgcca | 1560 |
| Db | 1501 | CTCCAGTTTCAGCGGCTCTCTGCCGGCCTCTTCCTCTACATCACTCGGAGSAGCGCCA | 1560 |
| Qy | 1561 | ggccgtcggaagcagtatccgatatctccgacgcgcgtccaaagggaacacctctctcg | 1620 |
| Db | 1561 | GGCGTCGCGAAGCAGTACCGATCTCCGACGCGCGCTCAAAGGGCAACACTTCTCTCG | 1620 |
| Qy | 1621 | catctcggcgtcatcgggaactcgaccttgtgtgcttccccgtctcttcccagtt | 1680 |
| Db | 1621 | CATCTCGGCGCTCATCGGAGCTCGACCTTCGTGTGCTGCTCCCGTCTCTTCCGAGT | 1680 |
| Qy | 1681 | attccgtgacttccgcttgcgcgcagctgcccgacctactggaccgcgagcgcttcg | 1740 |
| Db | 1681 | ATTCCGTGTACTTCGCGCTTGCCCGCAGCTGCCGACCTACTTGACCCGCCGAGCGGTCTG | 1740 |
| Qy | 1741 | gctcgtccgccacaaggctcttcgactacgcgcgcggtctcaaccacgcgaccgacaact | 1800 |
| Db | 1741 | GCTCGTCGCCCAACAAAGGCGCTCTTCGACTACGCGCGCGCTCACCCACGCGACCGACAAC | 1800 |
| Qy | 1801 | cgtactacatcggtccactgtgaaaggcaagaagtgcgtctcgtccgtccagttccttcg | 1860 |
| Db | 1801 | CGTACTACATCGGCTCCATCTGGACGCGCAAGATGGGTCTCTGTCGCTCGAGTCCCTTCG | 1860 |
| Qy | 1861 | acggcgcgtcgcgcggttcatcgagacttcaaccgcgaacaaacacgctgccacaaga | 1920 |
| Db | 1861 | ACGGCGGCTCGCGGGCTTCATCGACGTTCAACCCGAACAACAACGCTGCCAACAGA | 1920 |
| Qy | 1921 | ccataacccttactggcgagcttcgactcgcgggcaagpagctctcttcaaacacgacga | 1980 |
| Db | 1921 | CCATCAACCTTACTTGGCCGACCTTCGACTCGCGGCAAGCAGCTCTCTTCAACACGACGA | 1980 |
| Qy | 1981 | cgaaggacacctctctccgcgcgacccgcgcgcatcgttgagacttcaagcttgaccact | 2040 |
| Db | 1981 | CGAGGGACACCCCTCTCTCCCGCGACCCCGGCATGTTGAGACTTCAAGCTTTCACCGACT | 2040 |
| Qy | 2041 | tggcgcagcgcgaagaccaaagtgcgactcttggcgtgggttcaactctcgttgacgcgg | 2100 |
| Db | 2041 | TGGCAGCGACCAAGAACCAAGTGCGACTTCTGGCGTGGTCAATCTCGGTGAACGCGG | 2100 |
| Qy | 2101 | gtctctaggcgctttctctccgactctcttctgttcttcgttgtttattcttcagttc | 2160 |
| Db | 2101 | GTCTTAGGGCTCTTCTCTCCGACTCTCTGCTCTTTCTGTTGTTATTTCTTCGATTC | 2160 |

Qy 2161 cgttgatcagccattcgtgtagctcactcagatagatagcagttggcgaagtgcgaaa 2220
|||||
Db 2161 cgttgatcagccattcgtgtagctcactcagatagatagcagttggcgaagtgcgaaa 2220
|||||

RESULT 2
AX268013
LOCUS AX268013 2220 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 3 from Patent WO0166767.
ACCESSION AX268013
VERSION AX268013.1 GI:16516555
KEYWORDS
SOURCE Rhodosporidium toruloides.
ORGANISM Rhodosporidium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae;
Rhodosporidium.
1 (sites)
REFERENCE Chiang, S.J. and Basch, J.D.
AUTHORS Direct production of deacetylcephalosporin c
TITLE Patent: WO 0166767-A 3 13-SEP-2001;
JOURNAL BRISTOL-MYERS SQUIBB COMPANY (US)
FEATURES Location/Qualifiers
source 1..2220
/organism="Rhodosporidium toruloides"
/db_xref="taxon:5286"
BASE COUNT 389 a 814 c 554 g 463 t
ORIGIN

Query Match 100.0%; Score 2220; DB 6; Length 2220;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggaaccacccgaactctgcccgtttctggtcttctccttctcctgctgcgcacccatcgct 60
|||||
Db 1 GGATCCACCCGAACCTGTCCCGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 60
|||||

Qy 61 ttccgactcgcgcacatctgcttcttaacctcttcaacctcttcaacctcttcaacctcttcaacct 120
|||||
Db 61 TTCCCGACTCGCCGCAATGCTTAACTTCTTAACTTCTTAACTTCTTAACTTCTTAACTTCTTAACT 120
|||||

Qy 121 ccagctgcctttcctctcgaactcctcctcctcctcctcctcctcctcctcctcctcctcctcct 180
|||||
Db 121 CCAGCTCGCTTGCCTCTCCGACTCCTCTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 180
|||||

Qy 181 tccgctgcactcgcactcgcactcgcactcgcactcgcactcgcactcgcactcgcactcgcactcgc 240
|||||
Db 181 TCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 240
|||||

Qy 241 ctactggtgycggaatccgctacgcctcgcctcgcctcgcctcgcctcgcctcgcctcgcctcgcct 300
|||||
Db 241 CTACTGTTGGCGGGAATCCGCTACGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 300
|||||

Qy 301 cgcacacacagcagcctcgcacacagcagcctcgcacacagcagcctcgcacacagcagcagcagc 360
|||||
Db 301 CGCAGCGCACAGGCGCTCGCAGCGCTCGCAGCGCTCGCAGCGCTCGCAGCGCTCGCAGCGCTCG 360
|||||

Qy 361 caggggaaccacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
|||||
Db 361 CGAGGGAACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 420
|||||

Qy 421 gcagaaacagcgtcgagagattgccttctcctcctcctcctcctcctcctcctcctcctcctcct 480
|||||
Db 421 GCAGAAACAGCGCTCGGAGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480
|||||

Qy 481 cagggggaacaaattctcccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 540
|||||
Db 481 CGAGGGGACAAATTCTCCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 540
|||||

Qy 541 tgcagacacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 600
|||||
Db 541 TGCAGACACCCGAGCGACTTTGCGGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 600
|||||

Qy 601 tgtaaatccagctacagctctcgcagcgtttggtttctcctcgtgccaagccatgaagga 660
|||||
Db 601 TGTAATCTCCAGTACCGTCTCGGCAGCTTTGGTTTCTCTCGCTGGCCAAAGCATGAAGGA 660
|||||

Qy 661 ctacgggtgtaacgaacgcccgttctgtagaccaggtgagtttcccgagatgatacccgccc 720
|||||
Db 661 CTACGGGTGTAACGAACGCCGCTTGTCTTGACCAAGTGTAGTTTCCCGCATGATACCCGCC 720
|||||

Qy 721 accttgcactcagtcagcgccttccgcgtcgcagcaattcgccttcaatgggttca 780
|||||
Db 721 ACCTTTCGACTCATGCTGACGCCCTCTCCGCTCGCAGCAATTGCGCCCTTCAATGGGTCA 780
|||||

Qy 781 acagcaogctcgaagttcggcggaaccccgatcaogttacgatttggggcgagtcgtc 840
|||||
Db 781 ACAGCACGTCGCAAGTTTCGGCGGCAACCCCGATCACGTTACGATTTTGGGCGGAGTCG 840
|||||

Qy 841 aggcaggggtccgttatgaaccagatcattgcaacgtgagccacccgaaccgatctcc 900
|||||
Db 841 AGGCAGGGGTCCGTTATGAACCAAGTATTTGCGAACGTGAGCCACCCGAACCGATCTCC 900
|||||

Qy 901 agcgaactttcc 960
|||||
Db 901 AGCGGACTTTCCCGCC 960
|||||

Qy 961 ccgtcaaggtctcgtgtctcaagaagccctcttccacgctgccaogctgctcctcgtct 1020
|||||
Db 961 CCGTCAAGGCTCTCGGTCTCAAGAAGCCCTCTTCCAGGCTGCGATCGGCTCTCTCCGCTCT 1020
|||||

Qy 1021 tctccctaccacgaagtaacactcctcctcgcagcagctcctcctcctcctcctcctcctcctc 1080
|||||
Db 1021 TCTCTCCCTTACCAAGCAAGTACAACTCCCTCTTCCGCGAGCTGCTTACTTCCCAACTCG 1080
|||||

Qy 1081 tctcggcgacaaactgcaccaaagcgcctcgtctcctcctcctcctcctcctcctcctcctc 1140
|||||
Db 1081 TCTCGGCGACAACTGCACCAAGCGCTCTGCTTCTTCTGCTTGTCTGAAGCTGTCGAGC 1140
|||||

Qy 1141 ctgcgcgctcgtcgcggcggtggaagaaactcggcggttcccgctcgggggttttgg 1200
|||||
Db 1141 CTGCGGCGCTCGCTCGCGGGCGGTGAAGAACTCGGCGCGTTCGCGTTCGGGTTTGGT 1200
|||||

Qy 1201 cgtatgcctcgtcgtcagcagcagcctcttgcagcagcagcagcagcagcagcagcagcag 1260
|||||
Db 1201 CGTATGCTCCCGTCTGACGGGACCTTCTTGAAGCTGAGCGCGCTGCTCTCTCTCTCGCA 1260
|||||

Qy 1261 aggcgaagaagaaactcaatgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1320
|||||
Db 1261 AGGCAAGAAGAACTCAATGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
|||||

Qy 1321 gacactgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1380
|||||
Db 1321 GACACTGCTGACCGGCTCGCAGAACTCTTCCCGGGATCAACAACTTCGACGAGATGA 1380
|||||

Qy 1381 gttccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1440
|||||
Db 1381 GTTCCGCTGACGCTCTGTTCCCGCAGGAGCTGAGTTGCTTCTTTCGGAAGATTAG 1440
|||||

Qy 1441 attcatactcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1500
|||||
Db 1441 ATTCAATTCAGTACGCCACTATTACAGACGACACGATGACGACGAGTTCGACGCGGT 1500
|||||

Qy 1501 ctccagttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1560
|||||
Db 1501 CTCCAGTTCAGCGGCTCTCTCGCGGCTCTTCCCTTACATCATCTCTCGGAGAGCGCA 1560
|||||

Qy 1561 ggcgctcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620
|||||
Db 1561 GGCGCTCGGAAGCAGTACCCGATCTCCGACGCGCGCTCAAGGGGCAACACTTCTCTCTCG 1620
|||||

Qy 1621 catctcggcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680
|||||
Db 1621 CATCTCGGCCCTCATCGGGACTCGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
|||||

| | | | | |
|---|---|------|--|------|
| D | b | 2101 | GTCCTACGGGCGTCTTCCCTCACCAGACTTCCTTTCTGTCTTCTCTTCTTATCTTGTCAGTTC | 2160 |
| Q | y | 2161 | cgttgatcgccattcgtgcgttagctactcagtagtagacgctggcaagtgcgaag | 2220 |
| D | b | 2161 | CATTGCATGCCCATTCGTGGGTAGCTCACTCGAATAAGACTTTGGCAAGTCCGAAA | 2220 |
| RESULT 4 AR034144 1738 bp DNA linear PAT 29-SEP-1999 | | | | |
| LOCUS | | | | |
| DEFINITION | | | | |
| SEQUENCE 1 from patent US 5869309. | | | | |
| ACCESSION | | | | |
| VERSION | | | | |
| KEYWORDS | | | | |
| SOURCE | | | | |
| ORGANISM | | | | |
| REFERENCE | | | | |
| AUTHORS | | | | |
| TITLE | | | | |
| JOURNAL | | | | |
| FEATURES | | | | |
| source | | | | |
| BASE COUNT | | | | |
| ORIGIN | | | | |
| Query Match 61.9%; Score 1375; DB 6; Length 1738; | | | | |
| Best Local Similarity 84.7%; Pred. No. 3.6e-203; | | | | |
| Matches 1738; Conservative 0; Mismatches 0; Indels 313; Gaps 5; | | | | |
| Q | y | 67 | actgcgcgatgctccttaacctttcacccctgcgcccttcctgcgtcgcagctcccagct | 125 |
| D | b | 1 | ACTCGCGCCATGCTCCTTAACCTCTTACCCCTCGCCTCCCTCGCTCGAGGCTCCAGCT | 60 |
| Q | y | 127 | cgcccttgcctctcgaccctccctctgcgcgcgcacgaaacccaacgagccccctccgt | 186 |
| D | b | 61 | CGCCTTTTGCCTCTCCGACCTCCCTCGTCGCGCGCACGAACCAACGAGCGCCCTCCGCT | 120 |
| Q | y | 187 | cgtcgacctcgctacgcccgcctaccaaggctactgaacgagaccgcggactctactg | 246 |
| D | b | 121 | CGTCGACCTCGGTACGCCCGCTACCAAGGCTACTTGAACGAGACCCCGGACTCTACTG | 180 |
| Q | y | 247 | gtggcgcggaatccgcctaegccctcggtctgcgcgcttcacaggctcctcagacgcgcgcgac | 306 |
| D | b | 181 | GTGGCGCGGAATCCGCTACGCCCTCGGCTCAGCGCTCCAGGCTCCTCAGACGCCCGCGAC | 240 |
| Q | y | 307 | gcacaaggccgctgcgcgaacgcactagtagtgacogatctgttggcgcgctagcgaggg | 366 |
| D | b | 241 | GCACAAGCGCTCGCAACGCGACTGATGTGGACCGATCTGTGTGGCGGCTAGCGAGGG | 300 |
| Q | y | 367 | aaccaaacacaccaaagggttgtgcgcgcctagcaacagctcgagcagcgccgcgcagaa | 426 |
| D | b | 301 | AACCAACACGACCAAGGGCTTGC CGCGCGCTAGCAACAGCTCGAGCAGCGCGCGCAGAA | 360 |
| Q | y | 427 | acaggcgtcgagagattgacctctctcaatgctgttgcgtcccgccgctcgtgcaggg | 486 |
| D | b | 361 | ACAGGCGTCGAGGATTGCGCTCTTCTCAATGTCTTCCCCCGCCGCGCTGTGTCGAGGG | 420 |
| Q | y | 487 | cgacaactctccgctcctcgtctacattcaccgaggtggctacgctctcggcgcatcgag | 546 |
| D | b | 421 | CGACAATCTTCCCCTCCTCGCTACATTCACGGAGGTGGCTACGCCCTTCGCGATCGGAG | 480 |
| Q | y | 547 | cacgcgcagcgaatttgcgcgccttcacccaagcacacgcgggaaccaaagatgctgttaa | 606 |
| D | b | 481 | CACCGCAGCGAGCTTTTGC CGCGCTTACCAAGACACACGGGAACCAAGAATGGTCTGTGTA | 540 |
| Q | y | 607 | tctccagtacgctcgcgcagcttggtttctcgtcgtgcgaagccatgaaggactcgg | 666 |
| D | b | 541 | TCTCCAGTAGCCGCTCGGCGAGCTTTGTTTTCTTCGCTGGCCACGACCATGAAGGACTACGG | 600 |
| Q | y | 667 | tgtaacacgaacgcgcgcttgcttgaccagaggtgagtttcccgcatcatatccgcgcaccttt | 726 |

```
Db 601 TGTAAAGAACGCGGCTTGTTCAGC----- 625
QY 727 cgactcatgctgacgctctcccgctcgagcaattcgcccttcaatgggttcaaacagca 786
Db 626 -----CAGCAATTGCGCCCTTCAATGGGTTCAACAGCA 657
QY 787 cgtctgaagtctcgagggaaccccgatcacgtttagatttggcgagctctcagggcg 846
Db 658 CGTCTGAAGTTGCGGGGCAACCCCGATCACGTTAGATTGGGGGAGGAGTCTCAGGGCG 717
QY 847 aggttcggttatgaacagatattcgaaactgagccacccgaacccgatctccagcgga 906
Db 718 AGGTCCGTTATGAACAGATCATTCGAAC----- 748
QY 907 ctttcccccccccccccccccgctgacctccctctgttcttgaggggggaacacggtca 966
Db 749 -----GGCGGCAACACCGTCA 764
QY 967 aggtctcggtctcaagaagccctcttccacgctgpcatcgctcctcgtcttctctcc 1026
Db 765 AGGCTCTCGTCTCAAGAACCCCTCTTCCACGCTGCCATCGGCTCCTCGCTCTCTCTCC 824
QY 1027 cctaccaagccaagtacaactcccccttcgagagctgtctctactcccaactcgtctcg 1086
Db 825 CTTACCAAGCAAGTACAACTCCCTCTCCGAGGCTGCTACTCCCAACTGCTCTCGG 884
QY 1087 cgacaaactcacaaagccgctgctctctgctgctcctgaagctgtgacgctgcgg 1146
Db 885 CGACAAACTGACAAAGCCGCTCGTCTGCTTTCGCTCGAAGCTGTGCAGCGTCCGG 944
QY 1147 cgtctgctgctggcggtgaagaactcggcggttcccgcttgcgggttttggctgatg 1206
Db 945 CGTCTGCTGGCGGGGCTGAAGACTCGCGCGGTTCCGTTTCGGGTTTGTGCTGATG 1004
QY 1207 tcccgctgctgacgggacctcttgaactgagcgcgctgctcctctctcgccaaaggca 1266
Db 1005 TCCGGTGTGTCGAGCGGACCTCTTACTAGCGCGCGCTCGCTCTTCTCGCAAGGGCA 1064
QY 1267 aagaacactcaatggcggtggtggagagcttctcgagtgcttcagatctcgtgacact 1326
Db 1065 AGAAGAACCTCAATGGC----- 1081
QY 1327 gtgacggctgctgacgacctcttcaacgggatcaacaacctcgacgaagatgagttccc 1386
Db 1082 -----AACCTCTACCGGGATCAACAACCTCGACGAAG----- 1115
QY 1387 gtgacggctgctgttgcgacgagagactgactgttcttcttgcgaagattagattcat 1446
Db 1116 -----GATTTCAT 1122
QY 1447 attcactgacccactattcagaacacagatcagcagcagctcgacgagctctccca 1506
Db 1123 ATTCACTGACCCACTATTGAGAACGACAGATCAGGACACAGTCGCGGCGGCTCTCCCA 1182
QY 1507 gtctgacgctcctcctcgccgctcttccctacatcacctcgagagagcagcagccgt 1566
Db 1183 GTTCGACCGCCTCTCTCGCGCCTTTCCTCCCTATACCTCGAGGAGCGCCAGGCCGT 1242
QY 1567 cgcgaagcagtaacccgattctcgacgcgccgctcaaaaggggaacacacctctctcgactc 1626
Db 1243 CGGGAAGCAGTACCGGATCTCCGACGCGCGGTCAAGGGGAACACCTTCTCTCGCATCTC 1302
QY 1627 ggcgctatcgagactcgacctctgtgtcggttcccgctgcttcttcgagattccg 1686
Db 1303 GGCGGTATCGGAGCTCGACCTTCGT----- 1329
QY 1687 ctgacttcccgctgtcccgagctgcccgcactactgacgcgagcggttcggtcgct 1746
Db 1330 -----CTGCCCGACCTACTGACCGCGGAGCGGTTCCGGCTCGT 1367
QY 1747 ccgccccaaaggcctcttctgactaocgcccgtctaccacgacgacgaactcgtact 1806
```

```
Db 1368 CGCCCCACAAGGCGCTTTCGACTAGCGCGGCTCACCACGCGACCGACACTCGTACT 1427
QY 1807 acatcgctccatctgaaacggcaaaagtcgctctcgtccgctccagctcttcgagcg 1866
Db 1428 ACATCGGCTCCATCTGAGACGGCAAGATCGGTCTCGTCCGTCAGTCTTTCGAGCGG 1487
QY 1867 cgtcggcggttcacatcgagacgttcaacccgaaacaaacacgctgcaacaagaccatca 1926
Db 1488 CGCTCGGCGGCTTCATCGAGACGTTCAACCCGAACAACAACGCTGCCAACAAGACCATCA 1547
QY 1927 acccttactggcgaagctcgactcggggaagcagctcctcttcaacacgacgacgaggg 1986
Db 1548 ACCCTTACTGCGCGAGCTTCGACTCGGCAAGCAGCTCTTTCACACACGACGACGAGG 1607
QY 1987 acacctctctcccgaccccgcatctgtgagacttcaagcttgaacgactttggca 2046
Db 1608 ACACCTCTCTCCCGCGGACCCCGGATCGTTGAGACTTCAAGCTTGACCGACTTTGGCA 1667
QY 2047 cgagcagaagacaaagtcgacttctggcggtgggtcaatctcgtgaaacgcggtctct 2106
Db 1668 CGAGCCAGAAAGACCAAGTGGGACTTCTGGGTGGGTCAATCTCGGTGAGCGGGTCTCT 1727
QY 2107 agcgctctttc 2117
Db 1728 AGCGCTCTTTC 1738

RESULT 5
BD005938
LOCUS BD005938 1738 bp DNA linear PAT 31-JAN-2002
DEFINITION Cephalosporin esterase gene from Rhodospiridium toruloides.
ACCESSION BD005938
VERSION BD005938.1 GI:18634309
KEYWORDS JP 2001501466-A/1.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1738)
AUTHORS Polifino, M., Tonzi, S. M., Usher, J. J., Burnett, W. V. and Romanc, G.
TITLES Cephalosporin esterase gene from Rhodospiridium toruloides
JOURNAL Patent: JP 2001501466-A 1 06-FEB-2001;
COMMENT BRISTOL MYERS SQUIBB CO
OS Unidentified
PN JP 2001501466-A/1
PD 06-FEB-2001
PF 11-SEP-1997 JP 1998514766
PR 18-SEP-1996 US 60/026929
PI MICHAEL POLIFINO, SEAN M TONZI, JOHN J USHER, WILLIAM V BURNETT,
PI GUNA ROMANCIK
PC C12P21/06, C12N1/00, C12N9/16, C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
source 1..1738
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 327 a 639 c 444 g 328 t
ORIGIN

Query Match 61.9%; Score 1375; DB 6; Length 1738;
Best Local Similarity 84.7%; Pred. No. 3.6e-203;
Matches 1738; Conservative 0; Mismatches 0; Indels 313; Gaps 5;

QY 67 actcgcgccatgctcttaacctttcaacctcttcaacctcgctcctcgtcgtcgagctccagct 126
Db 1 ACTCGCGCCATGCTCTTAACCTTCTTACCTCTTACCTCGCTCGCTCGCTCGCTCGCTCGCT 60
QY 127 cgcctttgctctccgacctctctcgtccgacgacgacgacgacgacgacgacgacgacg 186
Db 61 CGCCTTTCGCTCTCCGACCTCTCCCTCGTCCGCGCACCAACCAACGAGCCCTCCCT 120
```


QY 1627 ggcgtcatcgagactcgactcttcgtgtggttcccgctcgttcttcaggtattccg 1686
|||||
DB 1303 ggcgtcatcgagactcgactcttcgtgtggttcccgctcgttcttcaggtattccg 1329
|||||
QY 1687 ctgacttcccgcttgcgcgagctgcccagcttactgagccgagcggttcgctcgt 1746
|||||
DB 1330 -----CTGCCGACCTACTTGACCCCGGAGCGTTCGGCTCGT 1367
|||||
QY 1747 ccgcccagaaggcctcttcgactacgcgcggctcaccagcgagccgacacactcgtact 1806
|||||
DB 1368 CCGCCCAAGGGGCTCTTCGACTACGCGCGGCTCACCACGCGACGACAACTCGTACT 1427
|||||
QY 1807 acatcggtccatcgaaagcggaagtggtctcgttcggttcggttcggttcgagcgcg 1866
|||||
DB 1428 ACATCGGCTCCATCGAACGGCAAGAGTGGTCTCGTCCGTCAGTCCCTTCGACGGCG 1487
|||||
QY 1667 cgtcgcggcttcacgagacgttcaaccggaacaaacagcgtgccaaagaccatca 1926
|||||
DB 1488 CGCTCGGGGCTTCATCGAGACGTTCAACCGGAACAACAACGCTGCCAACAGACCTCA 1547
|||||
QY 1927 accctactgcccagcttgcagtcgagcagcagctccttcaaacacgacgacgaggg 1986
|||||
DB 1348 ACCCTTACTGCGGACGTTGACTCGGCAAGCAGCTCTCTTCAACACGACGACGAGGG 1607
|||||
QY 1987 aacccctcttcgcgcgaccccgagtcggttgagcttcaagcttgcagactttggca 2046
|||||
DB 1608 ACACCCCTCTTCCCGCGGACCGCGCATCGTTGAGACTTCAAGCTTCAACGACTTTGGCA 1667
|||||
QY 2047 cgagccagagacaaagtgcgactctgggtgggtggttcaatcctgtaacggtctct 2106
|||||
DB 1668 CGACCCAGAACCAAGTGCAGCTTCGGCATGGGTCAATCTCGGTGAACCGGGTCTCT 1727
|||||
QY 2107 agcgctcttc 2117
|||||
DB 1728 AGCGCTCTTC 1738
|||||

RESULT 7

AX268011
LOCUS AX268011 1716 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 1 from Patent WO0166767.
ACCESSION AX268011
VERSION AX268011.1 GI:16516554

KEYWORDS Rhodospiridium toruloides.

SOURCE

Rhodospiridium toruloides.
Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae;
Rhodospiridium.

REFERENCE

1. (sites)
AUTHORS Chiang, S.J. and Basch, J.D.
TITLE Direct production of deacetylcephalosporin c
JOURNAL Patent: WO 0166767-A 1 13-SEP-2001;
BRISTOL-MYERS SQUIBB COMPANY (US)

FEATURES

source
1. 1716
/organism="Rhodospiridium toruloides"
/db_xref="taxon:5286"
BASE COUNT 325 a 630 c 439 g 322 t
ORIGIN

Query Match 60.9%; Score 1353; DB 6; Length 1716;
Best Local Similarity 84.6%; Pred. No. 9.1e-200;
Matches 1716; Conservative 0; Mismatches 0; Indels 313; Gaps 5;

QY 77 atgctcttaacctcttcacccctgcctccctcgttcgttcgagctccagctcgttcgccc 136
|||||
DB 1 ATGCTCTTAACCTCTTCAACCTCGCTCGCTCGCTCGAGGCTCCAGCTCGCCTTTGCC 60
|||||
QY 137 tctcagacctccctcgtcccgccgacgaacccaaacagagcccccctccctcgttcgacctc 196
|||||
DB 61 TCTCCGACCTCCCTTCGTCGCCCGCACGAACCAACAGAGCCCTCCCTCCGTCGTCGACCTC 120
|||||

QY 197 ggcctacgcccgcctaccaggctacttgaacgagaccgcccgaactctactgtgctcgga 256
|||||
DB 121 GGCTACGCCGCTACCAAGGCTACTTGAACGAGACCGCGGACTCTACTGTGTGGCGGGA 180
|||||
QY 257 atcgcgtacgctcggctcagcgcttccaggtcctcagaagcccgagcgacacaaagcc 316
|||||
DB 181 ATCCGCTACGCGCTCGGCTCAGCGCTTCAGGCTCTCAGACGCGCGGACCAAGGCC 240
|||||
QY 317 gtcgcgaacgagctagatggaccgatctgttggccggttagcgaggggaacaaacag 376
|||||
DB 241 GTCCGCAACGCGACTAGTATGGACCGATCTGTGGCCGGCTACGAGGGAACCAACAG 300
|||||
QY 377 accaagggttcgcccgcgcctagcaacagctcgcagcagcgccgcgagaaacagcgctcg 436
|||||
DB 301 ACCAAGGCTTGGCGCGCTTAGCAACAGCTCAGACGCGCGCGCAGAAACAGCGCTCG 360
|||||
QY 437 gagattcgctcttctcctcaatgctgttgcctcccgccggctcgtgcgagggcgcaaatctt 496
|||||
DB 361 GAGATTGGCTCTTCTCAATGTCTTGGCCCGCGCGCTCGTGGAGGGGACAAATCTT 420
|||||
QY 497 cccgtcctcgtctacattcacgaggttggctacgcttccgctcgtcgcgagcacgcggcag 556
|||||
DB 421 CCCGTCTCTGTACATTCACGAGGTGGCTTACGCCCTTCGGCGGATGCGAGCACCGGACG 480
|||||
QY 557 gactttcgccgcttcaacaaagcacacggaacaaagatggtcgttgaataatctccagtac 616
|||||
DB 481 GACTTTGGCGCTTACCAAGCACACAGCAACCAAGATGGTCTGTGTAATCTCCAGTAC 540
|||||
QY 617 cgtcggcagcttggcttctcgtcggcgaacccatgaaggactacggttgaacgaac 676
|||||
DB 541 CGTCTCGGACGCTTGTCTTCTCGCTGGCCAAAGCCATGAAGGACTACGGTGTAAACGAAC 600
|||||
QY 677 gcggcttgcctgacccaggtgagttcccgcatgatacccgccacacttccgactcagc 736
|||||
DB 601 GCCGCTTGTGAC----- 615
|||||
QY 737 tgacgctctcccgctcgcagcaattcgcccttcaatgggttcaacagcagctctcgaag 796
|||||
DB 616 -----CAGCAATTGGCTTCAATGGGTTCACACAGCAGCTCTCGAAG 657
|||||
QY 797 ttcggcggaaccccgctacgcttaccgatttggggcgagctgcagggcgagggctcgtt 856
|||||
DB 658 TTCGGCGCAACCCCGCATACGTTACGATTTCGGGCGAGTCTGCAGCGCAGGTCCTGTT 717
|||||
QY 857 atgaaccagatcattgcgaacgtgagccacccgaacgatctccagcgcgacttcccccc 916
|||||
DB 718 ATGAACCATCATTTGGAAC----- 738
|||||
QY 917 ccccccctcccgctgacctcctcctcgttctgagggcggaacacccgctcaaggctctcg 976
|||||
DB 739 -----GGCGGCAACACCGCTCAAGGCTCTCGG 764
|||||
QY 977 tctcaagaagccctcttcaacgctgcacgtcgtcctcgttcttctccctaccacga 1036
|||||
DB 765 TCTCAAGAAGCCCTCTTCCACGCTGCATCGGCTCTCTTCTTCTCCCTTCCCTACCAAGC 824
|||||
QY 1037 caagtacaactcccccttcccgagctcctactcccaactcgtctcggcgacaaactg 1096
|||||
DB 825 CAAGTACAACCTCCCTTCGCGGAGCTCTACTCCCAACTCGTCTCGGCGCAAACTG 884
|||||
QY 1097 caccagaagcgctcctcctcgttgcctcgaagctgcagcgtcgcgctcgctgc 1156
|||||
DB 885 CACCAAGCGGCTCGCTTCGCTTGCCTGGAAGCTGTCGACGCTCGCGGCTCGCTGC 944
|||||
QY 1157 ggcggcgtaagaactcgcgcggcggttcccggttgcgggttttgcgtatgtcccggtcgt 1216
|||||
DB 945 GCGGGGCTGAAGAACTCGCGGGCTTCCCTCGGTTTGGTCTGATGTCTCCGGTCTGT 1004
|||||
QY 1217 cgaaggaccttcttgactgagcgcgctccttctcgcaggggcaagaacact 1276
|||||
DB 1005 CGACGGGACCTTCTTGACTGAGCGGGCTCGCTCTCTTCTCGCCCAAGGGCAAGAACCT 1064
|||||

QY 1277 caatggcgtgctggcgagcttttcagagtgcttcagagatctcgtgacactgtcgacgcgc 1336
|||||
Db 1065 CAATGGC----- 1071
QY 1337 tcgcagaaactcttcaccggaggaatacaacaactcgcagaaagatagttcccgctgcagcgct 1396
|||||
Db 1072 -----AACCTCTTCACCGGGATCAACAACTCGACGAAG----- 1105
QY 1397 ctgttgcgccagcgagactgactgttcttttgcgaagatcagcattcatattcactgac 1456
|||||
Db 1106 -----GATCATATTCACTGAC 1122
QY 1457 gccactattcagaacacagacagcagcagcagcagcagcagcagcagcagcagcagcagc 1516
|||||
Db 1123 GCCACTATTCAAGACGACAGATCAGGACAGTCGCGAGCGCTCCCGAGTTCGACCGC 1182
QY 1517 ctctgcgcgcgtcttccctccatcactacacactcgcagagagcagcagcagcagcagcagcag 1576
|||||
Db 1183 CTCTCGCCGGCCTCTTCCCTTACATCACCCTCGAGAGCGCCAGGCGCTCGCGAAGCAG 1242
QY 1577 taccgatctccgacgcgcgtcaagggaacacacctctctcgcacatcgcgcgtcatc 1636
|||||
Db 1243 TACCGATCTCCGACGCGCGCTCAAGGGCAACACCTCTCTCGCATCTCGCGCGTCACTC 1302
QY 1637 gcggactcgacctctgtgtgcttcccccgtctcttctccgagttattccgctgactccc 1696
|||||
Db 1303 GCGGACTCGACCTCGT----- 1319
QY 1697 gcttgcccgagctgcccgaactactgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1756
|||||
Db 1320 -----CTGCCGACCTACTGACCGCGGAGCGTTCGGCTCGTCCGCCACAA 1367
QY 1757 gggcctctgactacgc 1816
|||||
Db 1368 GGGCTCTTGCATACGCGCGCTCACACGCGGACGACCACTCGTACTACATCGGCTC 1427
QY 1817 catctgaacgcgaagtcggtctctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 1876
|||||
Db 1428 CATCTGAACGGCAAGATCGGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCG 1487
QY 1877 ctctcagagacgttcaaccggaacaaacagctgcgaacagacacacacacacacacacac 1936
|||||
Db 1488 CTTTCATGACGCTTCAACCGCAACAAACGCTGCGCAACGACCAACCAACCAACCAACCAAC 1547
QY 1937 gccgacttgcactcggc 1996
|||||
Db 1548 GCGGACGCTTGCATCGCGGCAAGAGCTCTCTTCAACACGACGACGAGGACACCCCTCTC 1607
QY 1997 tccgc 2056
|||||
Db 1608 TCCGCGGACCGCGGCATCGTTGAGACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 1667
QY 2057 gaccaatgcgactctcggcgtgggtcaactcgcgtgaacgcgcgcgcgcgcgcgcgcgc 2105
|||||
Db 1668 GACCAAGTGCACCTCTGCGTGGGTGATCAATCTCGTGAACGCGGGGTCTC 1716
|||||
RESULT 8
AC087563/c
LOCUS 298166 bp DNA linear HTG 10-JAN-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-358F6, WORKING DRAFT
SEQUENCE, 75 unordered pieces.
ACCESSION AC087563
VERSION AC087563.1 GI:12061429
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 298166)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

2 (bases 1 to 298166)
DOE Joint Genome Institute.
Direct Submission
Submitted (10-JAN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 0
Center clone name: RPCI-11_358F6

Summary Statistics
Consensus quality: 189570 bases at least Q40
Consensus quality: 207567 bases at least Q30
Consensus quality: 217222 bases at least Q20
Estimated insert size: 231000; agarose-fp estimation
Estimated insert size: 290766; sum-of-contigs estimation
Quality coverage: 8.11 in Q20 bases; agarose-fp estimation
Quality coverage: 6.44 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1136: contig of 1136 bp in length
1137: gap of unknown length
1237: contig of 1021 bp in length
2258: gap of unknown length
2357: gap of unknown length
3495: contig of 1138 bp in length
3595: gap of unknown length
4702: contig of 1107 bp in length
4802: gap of unknown length
5944: contig of 1142 bp in length
6044: gap of unknown length
7153: contig of 1109 bp in length
7253: gap of unknown length
8400: contig of 1147 bp in length
8500: gap of unknown length
9505: contig of 1005 bp in length
9605: gap of unknown length
10744: contig of 1139 bp in length
10844: gap of unknown length
12291: contig of 1447 bp in length
12391: gap of unknown length
13519: contig of 1128 bp in length
13619: gap of unknown length
15101: contig of 1482 bp in length
15201: gap of unknown length
16501: contig of 1300 bp in length
16601: gap of unknown length
17767: contig of 1166 bp in length
17867: gap of unknown length
19234: contig of 1367 bp in length
19334: gap of unknown length
20427: contig of 1093 bp in length
20527: gap of unknown length
21623: contig of 1096 bp in length
21723: gap of unknown length
23056: contig of 1333 bp in length
23156: gap of unknown length
24305: contig of 1149 bp in length
24405: gap of unknown length
25647: contig of 1242 bp in length
25747: gap of unknown length
27088: contig of 1321 bp in length
27168: gap of unknown length
28319: contig of 1151 bp in length

QY 459 tcgttgcctccgcgcgtctgtagaggagacaattctccgtctctctgtatattcaag 518
|||||
Db 18145 NNNNNCCCCCCCCNN 18086
QY 519 gagtggtctagcctctggcgatgtagagacagcgagcacttgcgccttaaccaagc 578
|||||
Db 18085 NNNNNNNCCCCCCCCNN 18026
QY 579 acaggggaacaaagatggtgtgtgtaaatctcagtagctctcgagcttttggttcc 638
|||||
Db 18025 NNNNNCCCCNNCCNCCCGGGNNNTTTNCCNNAANNNNNNNNNNNNNNNNNNNNNN 17966
QY 639 tcgttggcgaagcatgaagtaagtgtaacgaacgcgcggttcttgaccagtgga 698
|||||
Db 17965 NAGTGGCGCTATTCAAGGNNTTTGANGACNCTNNCNNAACGGCNCCTTACAAGCTGTTA 17906
QY 699 gtttcccgatataccgcgcacacttctgaactcatgtaacgcctctccgcctgcagc 758
|||||
Db 17905 AAANNNNCCNCCGCGCCNCCCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17846
QY 759 aattcgccctcaatgggttcaacagcacgtctcgaagttcggcgcaaccccgatcacg 818
|||||
Db 17845 NNN 17786
QY 819 ttacgattggggcgagctgcaggcgaggttcgttatgaaccagatcattgcgaagc 878
|||||
Db 17785 NNN 17726
QY 879 tgagccaccgaaccgatctccagcgaatttcccccgcgcgcgcgcgcgcgcgcgcgc 938
|||||
Db 17725 CCGCCNNCCCGCCCGCCNNNNCCCGCCNNNNCCCGCCNNNNCCCGCCCGCCCGCCCGCC 17666
QY 939 ctgctcttgagggggcaacacgcgtcaaggctcgtctcgaagagccctcttccac 998
|||||
Db 17665 CCCCCCCCCNNCCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 17606
QY 999 gctgcatacgtctcgtct 1058
|||||
Db 17605 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 17546
QY 1059 gactgtctactcccaactcgtctcggcgacaaactgcacaaagcgcctctctctc 1118
|||||
Db 17545 CCCCCCCCCCGCCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 17486
QY 1119 gcttgctcgaagctgcagctgcggcgctgcggcgctgcggcgctgcggcgctgcggcg 1178
|||||
Db 17485 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 17426
QY 1179 gcgttcccgctcgggttttggtgtatgtcccggtgctgcaggggacatttctgactgag 1238
|||||
Db 17425 CCGCCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 17366
QY 1239 cgcgctgcctctctctcgaagggcaagagaacactcaatggcgtgctggtgcgagctt 1298
|||||
Db 17365 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 17306
QY 1299 tcgagtgctcaggtatctgtagactgtgcagcggctgcagaaactctt-----cac 1353
|||||
Db 17305 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 17246
QY 1354 cgggatacaaacctcgaacgaagatgagttcccgtagcggctgtgttcgcccagcagaga 1413
|||||
Db 17245 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 17186
QY 1414 ctgactgtcttttggaagattacgattattcactgacgacacattatcagaacga 1473
|||||
Db 17185 NNN 17126
QY 1474 cagcatcagcagcagtcgagcgcgtctccagtttcagacgcctctctcgcgcgcctctt 1533
|||||
Db 17125 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 17066
QY 1534 cccctacatacctcgtgagagcgccagcgctgcggaagcagtagtaccgagctctccgacgc 1593

Db 17065 CCCCCCCCCNN 17006
|||||
QY 1594 gcggtcaaaagggaacacacttctctgcgcatctgcggcgtcatcgcgactcgcacttcgt 1653
|||||
Db 17005 NCCNN 16946
QY 1654 gtgcgttcccccgtctcttctccgagttatccgctgacttcccgcttgcgcgagctgccc 1713
|||||
Db 16945 CCGCCCGCCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 16886
QY 1714 cgacctactgacgcgcgagcgttcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1773
|||||
Db 16885 NNN 16826
QY 1774 cgcgggtcaccagcgacgcgacacactcgtactacatcgcgtctccatctggaacggcgaaga 1833
|||||
Db 16825 CCGCCNN 16766
QY 1834 agtcggtctcgtcc--gtccagtccttcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1891
|||||
Db 16765 NNN 16706
QY 1892 caacccgaacaaacgcgtcgaacaaagacacacacacacacacacacacacacacacacac 1951
|||||
Db 16705 NATCAGTAGNTTANCAATTCCGGGNAATAACTTTCACCTNTNNNNNNNNNNNNNNNNNN 16646
QY 1952 gggcaagcagct 1963
Db 16645 NNNNNCCNCCCT 16634

RESULT 9

AC087190/c
LOCUS
DEFINITION
AC087190
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

165556 bp DNA linear HTG 25-APR-2001
Homo sapiens chromosome 16 clone RP11-47311, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
AC087190.2 GI:13786429
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 165556)
DOE Joint Genome Institute.
Direct Submission
Submitted (13-DEC-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 25, 2001 this sequence version replaced gi:11693350.

-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center clone name: RPCI-11-47311

Summary Statistics

Consensus quality: 112073 bases at least Q40
Consensus quality: 126903 bases at least Q30
Consensus quality: 135461 bases at least Q20
Estimated insert size: 138300; agarose-fp estimation
Quality coverage: 5.37 in Q20 bases; agarose-fp estimation
Quality coverage: 4.56 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces


```
* 60548 60647: gap of unknown length
* 60648 61743: Contig of 1096 bp in length
* 61744 61843: gap of unknown length
* 61844 62998: Contig of 1155 bp in length
* 62999 63098: gap of unknown length
* 63099 64220: Contig of 1122 bp in length
* 64221 64320: gap of unknown length
* 64321 65398: Contig of 1078 bp in length
* 65399 65498: gap of unknown length
* 65499 66762: Contig of 1264 bp in length
* 66763 66862: gap of unknown length
* 66863 68025: Contig of 1163 bp in length
* 68026 68125: gap of unknown length
* 68126 69400: Contig of 1275 bp in length
* 69401 69500: gap of unknown length
* 69501 70718: Contig of 1218 bp in length
* 70719 70818: gap of unknown length
* 70819 71995: Contig of 1177 bp in length
* 71996 72096: gap of unknown length
* 72097 73232: Contig of 1137 bp in length
* 73233 73333: gap of unknown length
* 73334 74476: Contig of 1144 bp in length
* 74477 74576: gap of unknown length
* 74577 75863: Contig of 1287 bp in length
* 75864 77021: Contig of 1058 bp in length
* 77022 77122: gap of unknown length
* 77123 78212: Contig of 1091 bp in length
* 78213 78312: gap of unknown length
* 78313 79424: Contig of 1112 bp in length
* 79425 79524: gap of unknown length
* 79525 80617: Contig of 1093 bp in length
* 80618 80717: gap of unknown length
* 80718 81804: Contig of 1087 bp in length
* 81805 81904: gap of unknown length
* 81905 83007: Contig of 1103 bp in length
* 83008 83107: gap of unknown length
* 83108 84167: Contig of 1060 bp in length
* 84168 84267: gap of unknown length
* 84268 85347: Contig of 1080 bp in length
* 85348 85447: gap of unknown length
* 85448 86521: Contig of 1074 bp in length
* 86522 86622: gap of unknown length
* 86623 87744: Contig of 1123 bp in length
* 87745 87844: gap of unknown length
* 87845 88965: Contig of 1121 bp in length
* 88966 89065: gap of unknown length
* 89066 90176: Contig of 1111 bp in length
* 90177 90276: gap of unknown length
* 90277 91279: Contig of 1003 bp in length
* 91280 91379: gap of unknown length
* 91380 92563: Contig of 1184 bp in length
* 92564 92663: gap of unknown length
* 92664 93688: Contig of 1025 bp in length
* 93689 94934: gap of unknown length
* 94935 95034: Contig of 1146 bp in length
* 95035 96111: Contig of 1077 bp in length
* 96112 96212: gap of unknown length
* 96213 98352: Contig of 2141 bp in length
* 98353 98452: gap of unknown length
* 98453 99645: Contig of 1193 bp in length
* 99646 99745: gap of unknown length

Query Match 3.38; Score 73.4; DB 2; Length 303091;
Best Local Similarity 21.58; Pred. No. 0.013;
Matches 239; Conservative 0; Mismatches 872; Indels 2; Gaps 1;

Qy 5 ccaccgaactgtccgcttttctgtcttctctgtctgtcgcgccatgcgtcttcc 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58104 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 58045

Qy 65 cgactgcgcgcatgtctcttaactcttctacactctctcctcgtctcctcgtcgcgacgtctccag 124
```



```
* 28612 28711: gap of unknown length
* 28712 35514: contig of 6803 bp in length
* 35515 35614: gap of unknown length
* 35615 49317: contig of 13703 bp in length
* 49318 49417: gap of unknown length
* 49418 69186: contig of 19769 bp in length
* 69187 69286: gap of unknown length
* 69287 98146: contig of 28860 bp in length
* 98147 98246: gap of unknown length
* 98247 127092: contig of 28846 bp in length
* 127093 127193: gap of unknown length
* 127193 181988: contig of 54796 bp in length.
FEATURES
    source
        1. 181988
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-759W11"
            /clone_lib="RPC1 human BAC library 11"
BASE COUNT 53222 a 36221 c 36728 g 51237 t 4580 others
ORIGIN

Query Match      3.18; Score 69.8; DB 2; Length 181988;
Best Local Similarity 20.5%; Pred. No. 0.051;
Matches 231; Conservative 0; Mismatches 894; Indels 3; Gaps 1;

QY 5 ccacccgaactctccgcttttggctttcttcttcttcttcttcttcttcttcttcttcc 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3185 CCNNNCCGCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 3126

QY 65 cgactgcgcgactctcttaactcttcttcttcttcttcttcttcttcttcttcttcc 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3125 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 3066

QY 125 ctgcgcttgcctcttcgactcttcttcttcttcttcttcttcttcttcttcttcc 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3065 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 3005

QY 185 gtcgtgactctgcgtacgcgcgtacaaagctactgactgactgactgactgacttac 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3005 CNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2946

QY 245 tggtagcggaatcgcgtacgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgcgtcgcg 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2945 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2886

QY 305 acgcacaaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2885 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2826

QY 365 ggaaccaacagcaagggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2825 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 2766

QY 425 aaacaggcgtcggaggattgcctcttcttcttcttcttcttcttcttcttcttctgag 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2765 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2706

QY 485 ggcacaattcttcgctctctctctctctctctctctctctctctctctctctctctcg 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2705 NCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 2646

QY 545 agcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2645 CGNNNCGCTNNNNNGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2586

QY 605 aatctccagtagctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2585 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2526

QY 665 ggtgtaacgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 2525 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2466
QY 725 ttgcactcatgctgaagcctctccgctcgcagcaattcgccttcaatgggttcaacag 784
Db 2465 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2406
QY 785 cactctcgaagtgcggcgcaaccccgatcacgttaagatttggggcgagtcgcaggc 844
Db 2405 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2346
QY 845 gcagggtcgttatgaaccagatcatgcaacgtgagccaccgacccgacgatccagcc 904
Db 2345 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2286
QY 905 gactttcccccccccccccccccccccccccccccccccccccccccccccccccccc 964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2285 NNNNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2229
QY 965 caaggtctcgttcaagaagccctcttccacgtcgcgcgcgcgcgcgcgcgcgcgcgcgc 1024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2228 CCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2169
QY 1025 cccctacaagcgaagtacaaactcccccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2168 CCCCNNNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2109
QY 1085 ggcgacaaactgcacaaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2108 CNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2061

RESULT 13
LOCUS AP001552 150120 bp DNA linear PLN 16-JUN-2000
DEFINITION Oryza sativa genomic DNA, chromosome 6, PAC clone:P0029D06.
ACCESSION AP001552
VERSION AP001552.1 GI:7363267
KEYWORDS
SOURCE
    ORGANISM
        Oryza sativa (cultivar:Nipponbare) DNA, clone:P0029D06.
        Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
        Ehrhartoideae; Oryzaceae; Oryza.
    REFERENCE
        AUTHORS
            Sasaki,T., Matsumoto,T. and Yamamoto,K.
        TITLE
            Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
            clone:P0029D06
        JOURNAL
            Published Only in Database (2000) In press
        REFERENCE
            2 (bases 1 to 150120)
        AUTHORS
            Sasaki,T., Matsumoto,T. and Yamamoto,K.
        TITLE
            Direct Submission
        JOURNAL
            Submitted (29-MAR-2000) Takuji Sasaki, National Institute of
            Agricultural Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@affrc.go.jp,
            URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441,
            Fax:81-298-38-7468)
    COMMENT
        The orientation of the sequence is from SP6 to T7 of the PAC clone.
        Genes were predicted from the integrated results of the
        following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
        SplicePredictor (October 1998 version). The genomic sequence was
        searched against the non-redundant database NRP (PIR, SWISSPROT,
        GENEPEP, PDB) from MAF DNA bank and the cDNA sequence database at
        RGP. Protein similarities of the coding regions were searched
        against NRP with BLASTP2.0. ESTs represent the identified cDNA
        sequences using BLASTN2.0 with the corresponding DDBJ accession no.
        and RGP clone ID.
        This sequence of this clone has an overlap with P0541H01 clone,
        DDBJ:AP001389 at the 3' end.
        This clone ends at the position 26,826 of P0541H01. Detailed
        information on overlap and assembly quality together with
        annotation of this entry at
        http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.
```

| FEATURES | Location/Qualifiers | source |
|----------|---|--------|
| CDS | <p>1. 150120 /organism="Oryza sativa" /cultivar="Nipponbare" /db_xref="taxon:4530" /chromosome="6" /clone="P0029D08" complement(join(1887..2096,2316..2390)) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93012.1" /db_xref="GI:7363268" /translation="MIKPIATVTLGRSLCEGLAAGRGKGCCEANRVSGLCPEADVE RKVEVAAGIEGSRWCRGLGAEVLSGSRKAGLAWPAAGGELARHEA" join(2761..3848,5363..5544,5878..5959,6846..7125) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93013.1" /db_xref="GI:7363269" /translation="MARRKGKDLISLPDDILLHILSMVRYKEAVTAASVREWKHLH TKLPALSFMSVLGAQGSLSLSTOSRORVDSMARTLRRACGPDRTVYKRLCLAYKDY PMBCHVADFIALAAASSLGLFNCNPKNLNRDDAGPWSLHLPAATACLSSMESCWYSVR PPHVGFGSALKSLTFKDSFWLHPGYLQDTAPPSELEHLHISGCTLSGSIETISATM PRKHLRIADSVVSGTAAIAVLADELTTLRVSDHGGKPPSPSSHEMLCYETLFR ASPTKYSYELRAPKLRFVDRCYAKEVVDVAVGRHLSDDVLEPAGRLPCYNEAK REFLOMDECKLNDILOGIMPGRWKYVQSLPYPRANKVRWGQDDREBEVEVQWHPR MDMLTMAASAGPGGCTAAAGDGVYMAATARAGEDGNSDLGSGVYMEDUGIGAAARW RADRARREMRPGGRGRRGNEAVTGWARRRAERYSIIVTAGGTSARLRRPDAAR TSGSAHEPDDRGAGAEGBEGAFSGYKFYKON" join(8143..8539,8557..8908,9092..9284) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93014.1" /db_xref="GI:7363270" /translation="MVVDKRPVAVPLARRPASRCRRLLDRRYAEQLIPPRKGGAPAG EGPVSREASSAGAGDASHIRCPSCSPPTLLCKSGTPSGVGRGERESGIGEMKSM APTSVELLSAGSAGEVELAGSGPSLLTGKSEVAVAHGEGARGHRIQAAVSSAA TPGPDGHRVPPADYVTRGIRREAPASGPDVVCAMRRRGLAVAAAVPPPPRP TPHRRHLPSTNHRHRYPCWLRSRGKRAAAAEVEEQRAAETPLPASEVRENK SSDRAHGLRRRGQRQVPRAGHRTMAVTVGEH" join(11170..11233,13292..13386) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93015.1" /db_xref="GI:7363271" /translation="MNLQVHPKRGKGVANGSDDKGILDGSDETNLFETSTKDFNS GAGKRQD" join(13886..13953,14703..14914,14920..15114,15144..15314, 15374..15492) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93016.1" /db_xref="GI:7363272" /translation="MNPEDQPSPADPTSPRPLPGADYKSGGKGTMDRAPLYSGSRE RGDAGTWAPLAVAGRGCGWAGPTGCHGSKREGAASGLGSAAPSPAGERBQAT DGDGREDRRGRHRLAATRMGAARAKAREGEVLTVGRYATREITTINGDA ELQDTRLEEMRVRGREGVHGEREPWPTAAAMEGLGTRETTFADGGIRWRGRRAA SGGRSGKGGGNDRGHGRWRKEKGEGCEAL" join(19096..19172,20123..20329,20389..20395) /note="hypothetical protein" /codon_start=1 /protein_id="BAA93017.1" /db_xref="GI:7363273" /translation="MQIEPTMHQCLVQWQDKTEIIVPANRQCSRRRLTAAPTRRP DIPLPSPESRRRTPLAASAAAPSRPPRASPGRSHSISGSGSALRLSLGLRI" 21587..21769 /note="5' LTR" join(21843..22489,22572..23072,23168..24293) /note="Similar to Arabidopsis thaliana DNA chromosome 4, BAC clone F17A8; putative protein. (AL049482)" /codon_start=1 /protein_id="BAA93018.1" /db_xref="GI:7363274"</p> | LTR |
| CDS | <p>complement(join(29217..30334,31321..31609,31701..31862)) /note="3' LTR" /note="hypothetical protein" /codon_start=1 /protein_id="BAA93019.1" /db_xref="GI:7363275" /translation="MIQIKIKKKRDEVLVATRLGKVSLSQALRMDGRVSGDDADGE RPLRRRAQCQPKSKGLDMLGPMIITSLTHTPPRLASTPIGRAAAAAVATAGL VSPAPHGKLRVSKRSGTGAAGSGEDGHRRRRSPQRSPSLSAKGLMDMLHQLI IALSSFDQLLAFSGTCSRRAALSFPSTYTFPPHLKPDIPNSHLCSSFRVTL LYKCRWQDGPDKRTLSRCSAPQNTPNMRYLGCYGVLYESYENCLLVDMYTGAK VYKPLQSGANKETIYILTAPLNFISHLILCSRSIYFWGVGTHSEHPFGGRI LQVLEKGFEEFAMDFHRLTRMFAPQLSQEYVYVGMGEFVGHFKPWLVTISGDM LMDLSVGHSHSGTFOVFLDFSAQTAKKMKLEKLSALFSLDRNRNFTFCT NPERWGGKNCIYVAKPSDSDEPMTAVELQPIGATHCVYPYSHPLLRTEGCSOLE YLWVLPSTFINGVDQ" 32671..33849 /note="hypothetical protein" /codon_start=1 /protein_id="BAA93020.1" /db_xref="GI:7363276" /translation="MOSPSKSEASGAPATSVIEGMAELPGLLHISVALLGSLD LLAFTGCHSWRAAFSSYPSKSNFRILPPLVRPNVRYKAPYSSNGHRLRSCVID NDENFTPKCQIPQETLQRHMFAGSHGQICCRGYCLVVDVFTGAENVPPRLPSE NDEFYTGILTPITQNSHLLIISTOSLDFWPGSDSWSLKLPRNVQDILVFNFG QLIAVEYSLYTLQALPILRLEKIKTMDNMNECPYMPFVYVCGDMLLIVDHYISF SFGAPVLYRPLRDMTKPAKVVEKLENAWLFIGDARSPPFKNPRMWRGSGNC LYAHYSQPSLSLHGLGDADAVDNDNLVFRKRWYRQLQALWYVPSFYSGDQQ " complement(join(34717..34935,35401..35912,36242..36363, 36470..36542,36626..36754,36936..37247,37336..38617, 38724..39212)) /note="Similar to Arabidopsis thaliana DNA chromosome V, BAC clone TM021B04; N. tabacum membrane-associated salt-inducible protein. (AF007271)" /codon_start=1 /protein_id="BAA93021.1" /db_xref="GI:7363277" /translation="MAAATAGAAATAVSTCSYSDVDDGAATSWLSGTSSTSSSS SORSORVRELLHDEAORLRRERGOGSGAFTPRVRRTPDMARYVEDDRAGHVGRH VAAVAAARATASCSPSSADMRAMASFAKLTREKCEVLHELHGWRRDFEFAH KLQCEPVSVAITILLRQGVQKVLAEVTFLEMLQACEPDVACGILLCAIYAKW KGLNDMLMFAVRRDIVPSISVFNFMSSQKQLHGVKIVHLWQMLAEANVAPQF TYTVVIGSAKEGMELEAMDAFEMRRRFRPEATYSLILSCAKHGEALGLYD EMKVSIVPSNYSVASCALTYLYKNEDYSKALSIFSEMQNKIVPDEVYIGLVYIKG LGIVEDAORMFEEDKAGLLSDQTVYMAAQVHMVQNTYDRALQVLDAMRNVKPSQ FYSALLRCHVAKEDVDAEDTFRALSNYPDPVFCNDLLRLYRLHLDKARALIL MKREALQFDEDLCTVLEVCCTSKNKDNTLNTLQNESSKSVLNPTDSTLSMM KLKLLDPGLSSVSQLIMKFAREGSTDEAKFLYELHTELGAKPDPTATLIVQYGQ AQLQEQAKLFEATFSTPVGGVYNAMVDALCGKTEAYRFLMELIDQGHNGAV TISILYTLTKQKFOEAEINIVRCLHDEALDTVYNTFKSMLEGLDNVFLAVS YADRMSSGIPRSMOTENINISYVGOGKLEKAVEMPSAAQELGLPDKETYNMLSF YKAGLHNEAEIIFQEKQNNHVPDSTHTYLALRAVTEGKCYKAEAIOMLRSNWT PSCSTHNLHLSAFLEKQIDEAQMYNOMEAGIPADLACCRMKHMLDHGVDDGI LFTFACRLKPDPSFILSAFHLHYHSEHSEAGVLDIAINMSGASFRLNLKVGSKLE QIQPKWMSIVKSGPLFPKAGHVPKTKQKSPWCCSVQEVVAREGEVDAQAQREGSRA AVHDHVORYDSSTCD" complement(join(40023..40413,40774..40971,41292..41464,</p> | CDS |
| CDS | <p>complement(join(34717..34935,35401..35912,36242..36363, 36470..36542,36626..36754,36936..37247,37336..38617, 38724..39212)) /note="Similar to Arabidopsis thaliana DNA chromosome V, BAC clone TM021B04; N. tabacum membrane-associated salt-inducible protein. (AF007271)" /codon_start=1 /protein_id="BAA93021.1" /db_xref="GI:7363277" /translation="MAAATAGAAATAVSTCSYSDVDDGAATSWLSGTSSTSSSS SORSORVRELLHDEAORLRRERGOGSGAFTPRVRRTPDMARYVEDDRAGHVGRH VAAVAAARATASCSPSSADMRAMASFAKLTREKCEVLHELHGWRRDFEFAH KLQCEPVSVAITILLRQGVQKVLAEVTFLEMLQACEPDVACGILLCAIYAKW KGLNDMLMFAVRRDIVPSISVFNFMSSQKQLHGVKIVHLWQMLAEANVAPQF TYTVVIGSAKEGMELEAMDAFEMRRRFRPEATYSLILSCAKHGEALGLYD EMKVSIVPSNYSVASCALTYLYKNEDYSKALSIFSEMQNKIVPDEVYIGLVYIKG LGIVEDAORMFEEDKAGLLSDQTVYMAAQVHMVQNTYDRALQVLDAMRNVKPSQ FYSALLRCHVAKEDVDAEDTFRALSNYPDPVFCNDLLRLYRLHLDKARALIL MKREALQFDEDLCTVLEVCCTSKNKDNTLNTLQNESSKSVLNPTDSTLSMM KLKLLDPGLSSVSQLIMKFAREGSTDEAKFLYELHTELGAKPDPTATLIVQYGQ AQLQEQAKLFEATFSTPVGGVYNAMVDALCGKTEAYRFLMELIDQGHNGAV TISILYTLTKQKFOEAEINIVRCLHDEALDTVYNTFKSMLEGLDNVFLAVS YADRMSSGIPRSMOTENINISYVGOGKLEKAVEMPSAAQELGLPDKETYNMLSF YKAGLHNEAEIIFQEKQNNHVPDSTHTYLALRAVTEGKCYKAEAIOMLRSNWT PSCSTHNLHLSAFLEKQIDEAQMYNOMEAGIPADLACCRMKHMLDHGVDDGI LFTFACRLKPDPSFILSAFHLHYHSEHSEAGVLDIAINMSGASFRLNLKVGSKLE QIQPKWMSIVKSGPLFPKAGHVPKTKQKSPWCCSVQEVVAREGEVDAQAQREGSRA AVHDHVORYDSSTCD" complement(join(40023..40413,40774..40971,41292..41464,</p> | CDS |

AUTHORS Bertolini, M., Laramee, L., Thomas, D., Cygier, M., Schrag, J. and Vernet, T.
 TITLE Polymorphism in the lipase genes of *Geotrichum candidum* strains
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1635)
 AUTHORS Bertolini, M.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1993) Bertolini M., Biotechnology Research Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada

FEATURES
 source Location/Qualifiers
 1..1635
 /organism="Galactomyces geotrichum"
 /strain="NRRL Y-552"
 /db_xref="taxon:27317"
 <1..1635
 /codon_start=1
 /product="lipase"
 /protein_id="AAA03428.1"
 /db_xref="GI:408459"
 /translation="OAPTAVLNGNEVISGLEKVDTFKGIPIFADPPVGDLPFKHPQP
 FNGSYOGLKANDFSSACMOLDPCNAISLDKVVGLGKIIPDNLRLGPLYDMAOGSVSMN
 EDCLYLVNRPAGTKPDALPVMWYIYGAFVGSASYPGNGYKESYEMGQPVYFV
 SINRTGPIYFGGDAITAEQNTNAGLDQRGLWVSDNIANFGGDPDKVMIFGESA
 DASVAHQILVAYGGDNTYNGKQLFHSAILQSGGPLEYFDSTSVGPESAYSREFAQYAGC
 DASAGNETLACLRKSSDVLHSAQNSYDLKDLGLLPQFLGFPDPDGNIIIPDAAYE
 LYRSGRYAKVPYITGNOEDGTLAPVAINATTPHVKWLYICSEASDAPLDRLVLS
 LYPGSGEAGPFTGILNALTQFKRIAAFTDLDLQSPRRVMLNATKDVNRWTVLAT
 QLHLNVPFLGTGHSLLFQYVDLGPSSARYRYFISFANHHDPNVGTLNQWDMYTD
 AKEMLIQHIGNSMRIDDFRIGISNFESDVTFLG"
 BASE COUNT 326 a 449 c 391 g 469 t
 ORIGIN

Query Match 3.0%; Score 67.6; DB 8; Length 1635;
 Best Local Similarity 56.4%; Pred. No. 0.32;
 Matches 149; Conservative 0; Mismatches 109; Indels 6; Gaps 1;
 QY 437 gaggattgcctcttccaatgtgtgccccccgctcgtgcgagggcgacaattctt 496
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 307 GAGGACTGTCTTACCTTACGTTTCCGCCCGCTGGCACCAAGCCTGATGCTAAGCTT 366
 QY 497 cccgtctctgtctacatcagaggtggtgtaagccttcggc-----gatgcgagcacc 550
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 367 CCCGTCATGTTTGGATTACGGTGGTGCCTTTGTGTTTGTGTTCTTCTTCTTACCT 426
 QY 551 ggcagcgacttgcgccttcacacagcagcaggaaccaagatggtcgtgtaaatctc 610
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 427 GGTAAACGGCTACGTCAAGAGAGTGTGGAATGGGTGAGCCTGTGTGTTGTTTCCATC 486
 QY 611 cagtaccgtctcgagcgttggttctcgtggtgccaagcattgaaggactacgggtga 670
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 487 AACTACCGTACCGGCCCTTATGATTCTCTGGTGGTGTGATGCCATCACCCTGAGGGTAAC 546
 QY 671 acgaacccggcttgcgttgaccag 694
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 547 ACCAACGCTGCTGTCACGCCAG 570

Search completed: June 29, 2002, 15:19:58
 Job time: 21479 sec

| Result No. | Query | Score | Match Length | DB ID | Description |
|------------|-------|-------|--------------|-------|-------------|
| 1 | 1 | 100 | 100 | 1 | 1 |

| | |
|------------|--|
| RESULT | 1 |
| AX268011 | |
| LOCUS | linear |
| DEFINITION | Sequence 1 from Patent WO0166767. |
| ACCESSION | AX268011 |
| VERSION | GI:16516554 |
| KEYWORDS | Rhodospiridium toruloides. |
| SOURCE | Rhodospiridium torulooides |
| ORGANISM | Eukaryota; Fungi; Basidiomycota; Urediniomycetes; Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae; Rhodospiridium. |
| REFERENCE | 1 (sites) |
| AUTHORS | Chiang,S.J. and Basch,J.D. |
| TITLE | Direct production of deacetylcephalosporin c |
| JOURNAL | Patent: WO 0166767-A 1 13-SEP-2001; |
| FEATURES | BRISTOL-MYERS SQUIBB COMPANY (US) Location/Qualifiers 1..1716 /organism="Rhodospiridium torulooides" /db_xref="taxon:5286" |
| source | |
| BASE COUNT | 325 a 630 c 439 g 322 t |
| ORIGIN | |

```
Query Match      100.0%; Score 1716; DB 6; Length 1716;
Best Local Similarity 100.0%; Pred. No. 3.6e-235;
Matches 1716; Conservative 0; Mismatches 0; Indels 0;
```

0:

[illegible]

| | | | |
|----|------|--|------|
| Qy | 1021 | actgagcgcgcgtcgtctcttctcgcgaagcggaagaacctcaatgcaacctcttc | 1080 |
| | | | |
| Db | 1021 | ACTGAGCGCGTCGCTCTTCTCGCCAAGSGCAGAGAACCTCAATGCAACCTCTTC | 1080 |
| Qy | 1081 | acgggatcaacaactcgacaagaattatcttactgacgcacattatcagaacgac | 1140 |
| | | | |
| Db | 1081 | ACGGGATCAACAACTCTGACGAAGATTCATATTCACTGAGCCACTATTGAGAAGAC | 1140 |
| Qy | 1141 | acgatcagcgaccagtcgcagcgcctctcccagttccgaccctctctcgcgccttc | 1200 |
| | | | |
| Db | 1141 | ACGATCAGCGACCAAGTCGACGCGCTCTCCAGTTCGACCGCCTCTCGCGGCTCTTC | 1200 |
| Qy | 1201 | cctacatcaccttcgagagacgcagccgcctcgcaagcagtagtacccgatctccgacgcg | 1260 |
| | | | |
| Db | 1201 | CCCTACATCACTTCGGAGAGCGCCAGGCGCTCGGCAAGTAGTACCCGATCTCGACGCG | 1260 |
| Qy | 1261 | ccgtcaaaaggcacaacctctctcgcatactcggccgctcgcgcgactcgacacttcgct | 1320 |
| | | | |
| Db | 1261 | CCGTCAAAAGGCAACACCTTCTCTGCACTCTCGGCGCTCATCGGGACTCGACCTTCGTC | 1320 |
| Qy | 1321 | tgcgcgaactactggaacgcgcgagcgcttcggctcgtccgcccacaaaggcctcttcgac | 1380 |
| | | | |
| Db | 1321 | TGCCCGACCTTACTTGAGACCGCGGAGGCGTTTCGGCTCTCGCCACCAAGGGCTCTTCGAC | 1380 |
| Qy | 1381 | taecgcgcggctcaccacgcgcgcgacaaactcgtactacatcggctccatctggaacggc | 1440 |
| | | | |
| Db | 1381 | TACGCGCGCGCTCACCACGCGACCGACAACTCGTACTACATCGGCTCCATCTGGAAGGCG | 1440 |
| Qy | 1441 | aagaagtcggtctcgttcgtccagtcttcgagcggcgcgctcgcgagcttcacgagacg | 1500 |
| | | | |
| Db | 1441 | AAGAAGTCGGTCTGTCGCGTCCAGTCTCTGACGCGCGCGCTCGGCGGCTTCATCGAGACG | 1500 |
| Qy | 1501 | ttcaaccgcgaacaaacacgcgtgcgaacaagaccataccacttactggtccgacgttcgac | 1560 |
| | | | |
| Db | 1501 | TTCAACCCGAGAACAAACAGCTGCCAACAAAGACCATCAACCCCTTACTTGGCCGACGTTGCAC | 1560 |
| Qy | 1561 | tcgggcaagcagctctcttcaacacgacgacgagggacacacctctctccgcgcgacccg | 1620 |
| | | | |
| Db | 1561 | TCGGGCAAGCAGCTCTCTTCAACACGACGACGAGGAGCACCTCTCTCCGCGGACCCG | 1620 |
| Qy | 1621 | cgcattcttagacttcaagcttgaccgactttggcacgacgcagaagaaccaagtgcgac | 1680 |
| | | | |
| Db | 1621 | CGATCTGTTGAGACTTCAAGCTTGCAGGACTTTGGCACGAGCACCAAGTGCAC | 1680 |
| Qy | 1681 | tctggcggtgggtcaatctcgttgaaacggggtctc | 1716 |
| Db | 1681 | TTCTGGCGTGGGTCAATCTCGGTGAACGGGGTCTC | 1716 |

| | | | | | | |
|--|------------|---|------------------------------------|---------|--------|-----------------|
| | RESULT | 2 | | | | PAT 29-SEP-1999 |
| | AR034144 | | AR034144 | DNA | linear | |
| | LOCUS | | Sequence 1 from patent US 5869309. | 1738 bp | | |
| | DEFINITION | | | | | |

| | | |
|-----------|------------|------------|
| ACCESSION | AR034144 | |
| VERSION | AR034144.1 | GI:5949749 |
| KEYWORDS | . | |
| SOURCE | Unknown. | |

| SOURCE | ORGANISM | UNKNOWN: |
|--------|----------|----------|
| | | Unknown: |

ORGANISM UNKNOWN.
Unclassified.

REFERENCE 1 (pages 1 to 1738)
Unclassified.

REFERENCE
1 (bases 1 to 1738)

AUTHORS Politino, M., Tonzi, S.M., Usher, J.J.

[illegible]

JOURNAL Patent: US 5869309-A 1 09-FEB-1999;

FEATURES

Location/Qualifiers

| | |
|-------------------|----------|
| FEATURES | 1. :1738 |
| LOCATION/QUANTITY | |

```

: source
: I. : I/38
: /organism="unknown"

```

| | | | |
|-----|---|-------------------------|-----|
| | : | : organism="unknown" | 328 |
| 297 | : | 629 | 444 |

| | | | | |
|------------|-------|-------|-------|-----|
| BASE COUNT | 327 a | 639 c | 444 g | 328 |
|------------|-------|-------|-------|-----|

ORIGIN

Query Match 100.0%; Score 1716;

Query Match
Best Local Similarity 100.0%
Best Local Similarity 100.0%
Pred. No. 3.

Best Local Similarity 100.0%; Pred. NO. 3.

\ddot{O} [illegible]

3.

[illegible]

[illegible]

Db 1757 GGGCTCTTCAGTACGCGCGGCTCAACACGCGACGACAACTCGTACTACATCGGGTC 1816
Qy 1428 catctggaacggaagagtgctgctcgtccagtcctctgacgagcgctcgggg 1487
Db 1817 CATCTGGAACGGCAAGAGTGGTCTCGTCCGTCAGTCCCTTCGACGGCGGCTCGGGG 1876
Qy 1498 ctteatcgagagcttaacccggaacaaacgctgcccgaacagaccattactg 1547
Db 1877 CTTTCATCAGAGCTTCAACCCGGAACACAGCGTGCACCAAGACCATCAACCCCTTACTG 1936
Qy 1548 gccagcttcaactcgggcaagagctcctcttcaacacgagcagagggagaccctctc 1607
Db 1937 GCCGACGCTTCGACGCGGCAAGAGCTCCTCTTCAACACGACGAGGAGACCCCTCTC 1996
Qy 1608 tccgcgcagccggcagctcgttgagacttcaagcttgaccgactttggcagcagcagaa 1667
Db 1997 TCCGCGCAGCCGGCATCGTTGAGACTTCAAGCTTGACCGACTTTGSCAGAGCCAGAA 2056
Qy 1668 gaccaagtgcagcttctgctggtggtcaatctcgtgtaacggtgtc 1716
Db 2057 GACCAAGTGCAGCTTCTGGCTGGGTCAATCTCGGTGAACGGGGTCTC 2105

RESULT 6

AX268013 AX268013 2220 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 3 from Patent WO0166767.

ACCESSION AX268013

VERSION AX268013.1 GI:16516555

KEYWORDS Rhodospiridium toruloides.

SOURCE Rhodospiridium toruloides.

ORGANISM Rhodospiridium toruloides.

REFERENCE Eukaryota; Fungi; Basidiomycota; Urediniomycetes;

Microbotryomycetidae; Heterogastridiales; Sporidiobolaceae;

Rhodospiridium.

1 (sites)

Chiang, S.-J. and Basch, J. D.

Direct production of deacetylcephalosporin c

Patent: WO 0166767-A 3 13-SEP-2001;

BRISTOL-MYERS SQUIBB COMPANY (US)

Location/Qualifiers

source

1. 2220

/organism="Rhodospiridium toruloides"

/db_xref="taxon:5286"

BASE COUNT 389 a 814 c 554 g 463 t

ORIGIN

Query Match 78.8%; Score 1353; DB 6; Length 2220;
Best Local Similarity 84.6%; Pred. No. 1.5e-183;
Matches 1716; Conservative 0; Mismatches 0; Indels 313; Gaps 5;

Qy 1 atgtccttaacctttcaacctgctcctcgtctgctgagagctccagctgctttgcc 60

Db 77 ATGCTCTTAACCTTTACCTGCTCCCTCCCTGCTGGAGCTCCAGCTCGCTTTGCC 136

Qy 61 tctccgacctcctcgtccgcgcagcaacccaaacagagccctcccgctgtagacctc 120

Db 137 TCTCGACCTCCCTGCTCGCGGCAGCAACCCAAACAGAGCCCTCCGCTCGACCTC 196

Qy 121 ggctacgcccgtaccaaggtacttgaacgagaccgcccgaactactgtggtggcga 180

Db 197 GGCTACGCCCGCTACCAAGGCTACTTGAACGAGACCGCGGACTCTACTGGTGGCGGA 256

Qy 181 atccgctacgctggtcagggcttccaggtcctcagagcccgccgacacaaagcc 240

Db 257 ATCCGCTACGGCTCGGGCTCAGCGCTTCAGGCTCCTCAAGCCCGCGACGACCAAGGCC 316

Qy 241 gtccgcaacgagctgagatggaccgactctgttggccggtctagcggaggaaccaacg 300

Db 317 GTCCGCAACGGGACTGAGTATGGACCGATCTGTGGCCGGCTAGCGAGGGAAACACAG 376

Qy 301 accaagggcttgccgccccttagcaacagctcagagcagcgccgcaaaaacagggctcg 360
Db 377 ACCAAGGGCTTGGCGCGCCTTAGCAACAGCTCGAGCAGCGCGCGCAAAACAGCGGTCG 436
Qy 361 gagatgctctctcctcaatgtggtgccccccgscggtcgtgagagggcgacaatttt 420
Db 437 GAGATTGCCCTCTCCCTCAATGTCTGTCGCCCGCGGCTCGTGAGAGGGGACAAATCTT 496
Qy 421 cccgtcctcgtctacattcacggaggtggctacgcttcctggtcgtatgagcaccggcagc 480
Db 497 CCCGTCTCTGCTACATTCACGGAGGTGGCTACGCTTCGGCGATGTCGAGCACGGCAGC 556
Qy 481 gacttgcgccccttcaacaacgaggaacaaagatggtcgttggtaaatctccagttac 540
Db 557 GACTTGGCGGCTTCACCAAGCACACGGGAACCAAGATGGTGGTGTAAATCTCCAGTAC 616
Qy 541 cgtctcgacagcttgggttctcctcgtgccaagcscatgaaggactcaggtgttaacgaac 600
Db 617 CGTCTCGGACACTTGGTTCCTCGCTGGCCAAAGCATGAAGGACTACGGTGTAAAGAAC 676
Qy 601 gccggttgcgttgac----- 615
Db 677 GCCGGCTTGTGACCAAGGTGAGTTTCCCGCATGATACCCGCCACCTTTCGACTCATGC 736
Qy 616 -----cagcaattcgccttcaatgggttcaacacagcagctctcgaag 657
Db 737 TGACGCCCTCTCCCGCTCGCAGCAATTCGCCCTTCAATGGGTTCAACACAGCTCTCGAAG 796
Qy 658 ttcggcgcaaaccccgatcacgtttacgatttggggcgagctctcagcgcgagggctccgtt 717
Db 797 TTCGGCGCAACCCCGATCACGTTACGATTTGGGGCGAGTCTGCAGGCGCAGGGTCCGTT 856
Qy 718 atgaaccagatcattggaac----- 738
Db 857 ATGAACAGATCATTTGCGAAGCTGAGCCACCGCAACCGATCTCCAGCGGACTTTCCCCCC 916
Qy 739 -----ggcgcaaacacggtcaaggctctcgg 764
Db 917 CCCCCCCCCCGCTGACCTCCCTCGTCTTCGAGGGCGGCAACACCGCTCAAGGCTCTCGG 976
Qy 765 tctcaagaagccctcttccagctgcatcggtcctcctcctcctccctccctcccaagc 824
Db 977 TCTCAAGAAGCCCTCTTCACGCTGCCATGGGCTCTCCGCTCTCTCCCTCCCTACCAAGC 1036
Qy 825 caagtacaactcctcctcgcgagctgctctactcctcaactcgtctcgtgagcaaaactg 884
Db 1037 CAAGTACAACCTCCCGCTTCGCCGAGCTGCTCTACTCCCAACTCGTCTCGGCGACAACATG 1096
Qy 885 caccaaagccgctcgtcctcgtcgtcgaagctgtagcagctgagcggtcgctcgtcgtc 944
Db 1097 CACCAAAAGCCGCTCGTCTTCGCTTGGCTCGAAGCTGTCGACGCTGCGGCGCTCGCTGC 1156
Qy 945 ggcggcgctgaagaactcggcggttcccgctcgggttttgggtggtatgtcccggtcgt 1004
Db 1157 GCGGGCGGTGAAGAACTCGGGCGGCTTCCCGTTCCGGGTTTGGTGTATGTCCCGGTGT 1216
Qy 1005 cgacgggaccttcttgaactgagcgcgctcgtcctctcgcgaagggaagaagaacct 1064
Db 1217 CGACGGGACCTTCTTGAAGTGAAGCGCGCTCGCTCTCTTCGCCAAGGGCAAGAAGACCT 1276
Qy 1065 caatggc----- 1071
Db 1277 CAATGGCGTGGCTGGCGAGCTTTCGAGTGTCTAGGATCTCGCTGACACTGTCGACCGGC 1336
Qy 1072 -----aacctcttcacgggatacaaacctcgcaggaag----- 1105
Db 1337 TCGGAGAACCCTTTCACCGGGATCAACAACTTCGAGAGAGATGAGTTCCTCGTCGAGGCT 1396
Qy 1106 -----gattcatattcactgac 1122
Db 1397 CTGTTGCCCGCAGCGAGACTGACTTGTCTCTTTTTCGGAAGATTACGATTATATTA 1456
Qy 1123 gccactattcagaacgacacgatcagcaccagtgagcagcggtctctccagttcagccgc 1182


```

Db 487 AACTACGTTACCGCCCTATGATCTCTGGTGGTATGCCATCACCGCTGAGGGTAAC 546
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 595 acgaagccgctgtgtgaccagcattcgcccttcaatgggttcaacagcagctctcg 654
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 547 ACCAACGCTGTGTCAGCAGCAGCAGCAGGCTCTGAGTGGTGTAGCGACAACATTGCC 606
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 655 aagttcggcggaaccccgatcagttacgatttggggcgagttctgcagcgcgagggctc 714
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 607 AACTTTGGTGTATCCGACAAGGTCATGATTTTCGGTGGTGGTGGTGGTGGTGGTGGT 666
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 715 gttatgaaccagatcattcgaaacgagcggaacacacgctcaa 755
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 667 GTTGCTCACCGCTGTGCTAGCTAGGTCAGGTCGACACACCTA 707
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 11
GCTAGL GCTAGL 1635 bp DNA linear PLN 25-JAN-1996
LOCUS G.candidum triacylglycerol lipase gene.
DEFINITION X81656
ACCESSION X81656
VERSION X81656.1 GI:547503
KEYWORDS triacylglycerol lipase.
SOURCE Galactomyces geotrichum.
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (bases 1 to 1635)
AUTHORS Pretorius, G.H.J.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1994) G.H.J. Pretorius, University of the Orange
Free States, Dept of Microbiology and Biochemistry, PO Box 339,
Bloemfontein, 9300, SOUTH AFRICA
REFERENCE 2 (bases 1 to 1635)
AUTHORS Phillips, A., Pretorius, G.H. and van Rensburg, H.G.
TITLE Molecular characterization of a Galactomyces geotrichum lipase,
another member of the cholinesterase/lipase family
JOURNAL Biochim. Biophys. Acta 1252 (2), 305-311 (1995)
MEDLINE 96049515
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="CBS 178.71"
/db_xref="taxon:27317"
<1. .1635
/EC_number="3.1.1.3"
/codon_start=1
/product="triacylglycerol lipase"
/protein_id="CAA57316.1"
/db_xref="GI:547504"
/translation="SPTREMBL:Q12614"
FTGSYGLKANDPSSACMOLDPNAISLDKVVGLGNILPDNLRGPLYDMAOGTSVMN
ECLLYNVRPAGTKPDALPVWVWLYGAGVFGSSASYPGNGYVKESVEMGPPVVF
SINRTGPGFLGGDAITAEAGNTNGLHQRKLEWSDNIANGFGDPKVMIFGESA
GAMSAVHQIAYGGDNTYNGKLFSAILOGSGPLFYFDSTVGPESAYSRFAQYAGC
DPSVADNETIACLRSSDLVLSAQNYSYDLKDLGLLPQFLGFGPRPDGNIIPDAAYD
LYRSGRYAKVPYITGNOEGDTILAPVAINATTPHVKWLYICSQASVSLDRVLS
LYPGSWSEGAFFRTGLNALTPOFKRIAAFTDQLFQSPRRVMLNATKDVNRWYTLAT
QLHNLVFFLGTGHSGLLFOYVVDLQSPSAYRYFFISFANHDPNVGTNLQWDMYTD
AGKMLQIHMVGNMRTDDFRIGISNFESDVILFG"
BASE COUNT 328 a 453 c 387 g 467 t
ORIGIN

Query Match 6.2%; Score 107; DB 8; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;
Matches 225; Conservative 0; Mismatches 170; Indels 6; Gaps 1;

QY 361 gaggattgctctctcaatgctgttgcggcgctgtgctgcgagggcgaactctt 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 307 GAGGACTGTCTCTACCTACAGCTTTCCGCCCGCTGGCACCAGCCTGGAGATACTC 366
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 421 ccggtctctgtacattcaaggaggtggtacgctctggc-----gatcgagcacc 474
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 CCCTCATGTTTGGATCTACGGTGGTCCCTTGTGTTGTTCTTCTGCTTCTTACCT 426
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 ggagcgacttgcgccttcaccaagcacacagcggaacccaagatggctgtgtaaatctc 534
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 GGTAAACGGCTACGTCRAGGAGAGTGTGAAATGGCCAGCCAGTTGTGTTTCCATC 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 cagtacgctcgcgcagcttgggttctcctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 594
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AACTACCGTACCCTCCCTATGATTTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 595 acgaagccgctgtgttgcagcagcattcgcccttcaatgggttcaacagcagctctcg 654
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 ACTAACTCTGCTCTGCTACGTCAGCGAAGGCTCTCAGTGGGTTCGCGACAACATTGCC 606
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 aagttcggcggaaccccgatcagttacgatttggggcgagttctgcagcgcgagggctc 714
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 AACTTTGGTGTATCCGACAAGGTCATGATTTTCGGTGGTGGTGGTGGTGGTGGTGGT 666
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 gttatgaaccagatcattcgaaacgagcggaacacacgctcaa 755
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 667 GTTGCTCACCGCTTATTGCTACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 707
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
GCU02387 GCU02387 1635 bp DNA linear PLN 08-OCT-1993
LOCUS Geotrichum candidum NRRL Y-553 lipase gene, partial cds.
DEFINITION U02387
ACCESSION U02387
VERSION U02387.1 GI:406507
KEYWORDS Galactomyces geotrichum.
SOURCE Galactomyces geotrichum
ORGANISM Galactomyces geotrichum
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
REFERENCE 1 (sites)
AUTHORS Bertolini, M., Laramée, L., Thomas, D., Cygler, M., Schrag, J. and
Vernet, T.
TITLE Polymorphism in the lipase genes of Geotrichum candidum strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1635)
AUTHORS Bertolini, M.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1993) Bertolini M., Biotechnology Research
Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2,
Canada
FEATURES
source Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="NRRL Y-553"
/db_xref="taxon:27317"
<1. .1635
/codon_start=1
/product="lipase"
/protein_id="AAA03425.1"
/db_xref="GI:406508"
/translation="QAPTAVLNGNEVISGLGKVDTFKGIPTADPPVGLRFRKHPQP
FTGSYGLKANDPSSACMOLDPNAISLDKVVGLGNILPDNLRGPLYDMAOGSVSMN
ECLLYNVRPAGTKPDALPVWVWLYGAGVFGSSASYPGNGYVKESVEMGPPVVF
SINRTGPGFLGGDAITAEAGNTNGLHQRKLEWSDNIANGFGDPKVMIFGESA
GAMSAVHQIAYGGDNTYNGKLFSAILOGSGPLFYFDSTVGPESAYSRFAQYAGC
DPSVADNETIACLRSSDLVLSAQNYSYDLKDLGLLPQFLGFGPRPDGNIIPDAAYE
LYRSGRYAKVPYITGNOEGDTILAPVAINATTPHVKWLYICSQASVSLDRVLS
LYPGSWSEGAFFRTGLNALTPOFKRIAAFTDQLFQSPRRVMLNATKDVNRWYTLAT
QLHNLVFFLGTGHSGLLFOYVVDLQSPSAYRYFFISFANHDPNVGTNLQWDMYTD
AGKMLQIHMVGNMRTDDFRIGISNFESDVILFG"
BASE COUNT 323 a 449 c 394 g 469 t
ORIGIN

Query Match 6.2%; Score 107; DB 8; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;

```


| | Matches | 225; Conservative | 0; Mismatches | 170; Indels | 6; Gaps | 1; |
|----|---------|--|---------------|-------------|---------|----|
| QY | 361 | gaggatggcctcttcctaagtgtggtgcccccgcgctcgtagagggcgacaattctt | 420 | | | |
| Db | 307 | GAGGACTGTCTTACCTTTACGTTTTCCGCCCTCGTGGCACAAGCCTGATGCTAAGTC | 366 | | | |
| QY | 421 | ccqgtccctgctacattcacggagggtggtacaccttcgc-----gatgcgagacc | 474 | | | |
| Db | 367 | CCCGTCATGGTTGGATTACGGTGGGCGCTTTGTTGGTTCTTCCTTCATTACCCT | 426 | | | |
| QY | 475 | ggcagcgactttgcgccttcaccaagcacacgaggaacaaagatggttggttaaatctc | 534 | | | |
| Db | 427 | GGTAACGGCTACGTCAGGAGAGTGTGGAAATGGGCCAGCCTGTTGTGTTTGTTCATC | 486 | | | |
| QY | 535 | cagtagcgtctcgagcgttgtgttctcgtcgccaagcatgaaggactacagggtga | 594 | | | |
| Db | 487 | AAC TACCGTACCGGCCCTCATGGATTCTCGGTGGTGNATGCCATCACCGCTAGGGGTAA C | 546 | | | |
| QY | 595 | acuaacgcccgtgtgttacccagcaattcgccctcaatgggttcaaacagcagctcgc | 654 | | | |
| Db | 547 | ACCAACGCTGGTCTGCAGCACCACGGAAGGCTCTCGAGTGGGTAGCGACAACATTGCC | 606 | | | |
| QY | 655 | aagttcggcgcaacccccgatcacgttacgattttggggcgagctgcaggcgccagggtcc | 714 | | | |
| Db | 607 | AAC TTGGTGGTGATCCGCACAAGGT CATGATTTTCGTGTAGTCCGCTGGTCCATGAGT | 666 | | | |
| QY | 715 | gttatgaaccagatcatctgcgaacggcggaacaccgtcaa | 755 | | | |
| Db | 667 | GT TGCTCACCAAGCTTGTTGCTTACGTTGGTGGTGACAAACCTA | 707 | | | |

| | | |
|------------|--------------------------------|---|
| RESULT | 13 | |
| LOCUS | GCU02541 | |
| LOCUS | 1635 bp | DNA linear |
| DEFINITION | Geotrichum candidum ATCC 34614 | lipase II gene, partial cds. |
| ACCESSION | U02541 | |
| VERSION | U02541.1 | GI:408458 |
| KEYWORDS | | |
| SOURCE | | Galactomyces geotrichum. |
| ORGANISM | | Galactomyces geotrichum. |
| | | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; |
| | | Saccharomycetales; Dipodascaceae; Galactomyces. |
| REFERENCE | | 1 (sites) |
| AUTHORS | | Bertolini, M., Laramee, L., Thomas, D., Cygler, M., Schrag, J. and Vernet, T. |
| TITLE | | Polymorphism in the lipase genes of Geotrichum candidum strains |
| JOURNAL | | Unpublished |
| REFERENCE | | 2 (bases 1 to 1635) |
| AUTHORS | | Bertolini, M. |
| TITLE | | Direct Submission |
| JOURNAL | | Submitted (14-OCT-1993) |
| | | Institute - NRCC, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada |

Location/Qualifiers
1. .1635
/organism="Galactomyces geotrichum"
/strain="ATCC 34614"
/db_xref="taxon:27317"
<1. .1635
/codon_start=1
/product="lipase II"
/protein_id="AAA03430.1"
/db_xref="GI:408469"
/translation="QAPTVLNGNEVIGSVGKVTDFKGIPEADPLNDLREKHQP
FVGSYGLKRNDESPACMDPENSITLDDAKGLAKVPIPEERGGDLYDMAKVTSMN
EGLYLKNDPARKFEDKLPVNWYIIIGAFVYGSAVPGNSIYKESINMGQVYFV
SINRTGPFGLGGDITAEQNTNAGLHQDKGLEWSDNIANGFGDDPKDKWYIFGESA
GAMSAVQHLIAGGDNTYNGKGLFHSAIIQSGDGLPLHDSSVGGPDISYNRNFAYAGC
DYSANDTECLRSKSSVLDHQAQSYDLKDLGLLPQFLGFGPDGNIINFDPAAYE
LRFSRYAKVYIISQNGDEGTAFAFVALNATTPHVKWLQYIFDYDASASDRVLS
LPQTLSVSGSPFTGLNALTQFKRVAAILSDMLEQSPRRVMSATKQVNRWTVYLT
HLHNVPTFLTGHNGLNIFQVNIIGPANSYLYRIFSFAHHDPNVGNTLLQWDQVTD
EKKEMLEIHTQNVMDTDYREGISNFTDNLVYG"

| BASE COUNT | 348 a | 460 c | 373 g | 454 t |
|------------|-------|-------|-------|-------|
| ORIGIN | | | | |

Query Match 6.2%; Score 107; DB 8; Length 1635;
Best Local Similarity 56.1%; Pred. No. 2.9e-06;
Matches 225: Conservative 0; Mismatches 170; Indels

| | | | | | |
|-----|----|---|---|---------------------|-----|
| 361 | Qy | gagatgagcctcttctc | caatgctggtgccccccgcgcgctcg | tcgcgagggcgacaatctt | 420 |
| | | | | | |
| 307 | Db | GAGACGTGCTTTACCTCAATGTTTCCGCCCTGCTGGACCAAGCCTGAAGATAAGCTC | 366 | | |
| | | | | | |
| 421 | Qy | cccgctcgtctctacattcaaggaggtgggtacgccttggc-----gatcgagcacc | 474 | | |
| | | | | | |
| 367 | Db | CCCGCTCAAGTGTGTGATTTACGGCGGTGCGTTGTGTACGGTTCCTGCTGCCTACCCCT | 426 | | |
| | | | | | |
| 475 | Qy | ggcagcgactttgcgccttcaccaaagcacacagggaaacaaagatggtctgttgaatatctc | 534 | | |
| | | | | | |
| 427 | Db | GGTAACAGCTACGTTAAGGAAGATATCAACATGGCGACGCCCGTTGTGTGTTCCATC | 486 | | |
| | | | | | |
| 535 | Qy | cagtaccgcttcgcgcagcttggttctctcgtcgtggcgaagcaatgaagactacggtgta | 594 | | |
| | | | | | |
| 487 | Db | AACCTAACCGTACC | GGTCCATTTGGATTCTCGTGGTGGTGATGCCATCACCGCTGAGGGCAAC | 546 | |
| | | | | | |
| 595 | Qy | acgaacgcgcgcgttgtagacaagcaattgcgcctcaatgggttcaacagcaacgtctcg | 654 | | |
| | | | | | |
| 547 | Db | ACCAACGCTGGTGTGCACGACCAAGCGGCTCGAGTGGGTAGCCACACACATTGGC | 606 | | |
| | | | | | |
| 655 | Qy | aagtctggcggcaaccccgatcagttacgatttggggcgagtcctcgagggcgagggcttc | 714 | | |
| | | | | | |
| 607 | Db | AACCTTGGTGGTGATCCCGACACAGGTCACTGATTTTGGTGAGTCCCGCTGGTGCCATGAGT | 666 | | |
| | | | | | |
| 715 | Qy | gttatgaaccagatcatcttgcgaacggcggaacacccgtcaa | 755 | | |
| | | | | | |
| 667 | Db | GTTGCTCACCAGCTTATTGCTTACGGTGGTGACAAACCTA | 707 | | |
| | | | | | |

RESULT 14

| | | | | |
|------------|---|-----------|--------|-----------------|
| GCU02622 | 1635 bp | DNA | linear | PLN 22-OCT-1993 |
| LOCUS | | | | |
| DEFINITION | Geotrichum candidum ATCC 34614 lipase gene, partial cds. | | | |
| ACCESSION | U02622 | | | |
| VERSION | U02622.1 | GI:409275 | | |
| KEYWORDS | | | | |
| SOURCE | Galactomyces geotrichum. | | | |
| ORGANISM | Galactomyces geotrichum | | | |
| | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| | Saccharomycetales; Dipodascaceae; Galactomyces. | | | |
| REFERENCE | 1 (sites) | | | |
| AUTHORS | Bertolini,M., Laramee,L., Thomas,D., Cygler,M., Schrag,J. and Vernet,J. | | | |
| TITLE | Polymorphism in the lipase genes of Geotrichum candidum strains | | | |
| JOURNAL | Unpublished | | | |
| REFERENCE | 2 (bases 1 to 1635) | | | |
| AUTHORS | Bertolini,M. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (21-OCT-1993), Bertolini M., Biotechnology Research | | | |
| | Institute-NRCC, 6100 Royalmount Avenue, Montreal, Quebec, H4P 2R2 | | | |
| | Canada | | | |

FEATURES
SOURCE

CDS

```
/usr_xref="gl:032576"/
/transl="0QAPTALVNGNEISGVLEGKVDTFGPADPVPVGDILRFKHQP
F*GSGYGLKANDESSACMLDPGNAISLLDKVVLGKGIIPDNLRGPLYDMAQGSVMN
EDCLLVNFRFAGTPDKALPVMTVYIGGAFVSSASPGNGKRSVSEVQPVYFV
SNYRTPGYPGLGDAITAGENTNAGLHPDQKGLFESWDNIANFGPDPKVMYIFGSA
```

GAMSAVLAHLVXGGDNTYNGKLFHSAILOQGGPLPFDSTSVGPESAYSRFAQYAGC
DASAGNETLACLRSDVHSAQNTDLKDLGLPQLFGPRDNGIIPDAAYE
LYSGRAGKAPVYITQNOEDGFIAPNAVNTYHPYKMLVTCSSADSLDRVLS
LYPSSWSEGAPEFTGLINALPOFRIAAITDILFOSPREVMNLAKFDVNRKTYLAT
QNLNLPVFLGTFHGSDLLFOYIVDLGPPSARVRFIFSAHHDDPVGCTNLKQDMXTD
SGKMLQIHMIGNISDQDTRFTEIGLNSFSDPVLFG#

BASE COUNT
ORIGIN

Query Match 6.2%; Score 106.8; DB 8; Length 1635;
Best Local Similarity 56.3%; Pred. No. 3.1e-06;
Matches 233; Conservative 0; Mismatches 167; Indels 6; Gaps 1;

| | | | |
|-----|----|---|-----|
| 361 | QY | gaggatgacctctctccaatgctgtgccccgcgcggtctgacgagggacaaattctt | 420 |
| | | | |
| 307 | Db | GAGGACTGTGCTCTACCTTAACGCTTTTCGCCGCCGTGGCCACCAAGCCTGATGCTAAGCTC | 366 |
| | | | |
| 421 | QY | ccggtcctgctctacattcaaggaggtgctacgctctggc-----gatcgagacacc | 474 |
| | | | |
| 367 | Db | CCCGTCATGTTTGGGATTACGGTGCTGCTTTGGTTCCTCTCTCTTTACCCCT | 426 |
| | | | |
| 475 | QY | ggcagcgacctttgcgccttccaccaagcacacgggaaccaagatggtctgttgaatatctc | 534 |
| | | | |
| 427 | Db | GGTAACGGCTACGTCAAGGAGAGGTGGAAATGGGCCACCGCTGTGTGTTCTTCCATC | 486 |
| | | | |
| 535 | QY | cagtaacctgtctggcagcttggtttctctcgtctggccaagccatgaaggaactacagtgta | 594 |
| | | | |
| 487 | Db | AACTAACGTACCGGCCCTATGGATTCCTGGGTGTGATGCCATCACCGCTGAGGGTAAC | 546 |
| | | | |
| 595 | QY | acgaacgcggctgtgtgacacgaattcgcccttcaatgggttcaacagcacgtctcg | 654 |
| | | | |
| 547 | Db | ACCAACGCTGGTCTGCACGACCAGCGCAAGGGTCTCGATGGGTGTAGCGACAACTATGCC | 606 |
| | | | |
| 655 | QY | aagtcgcgggcgaacccgatcacgttacgatttggcgagctctcagggcgacgggtcc | 714 |
| | | | |
| 607 | Db | AACCTTGGTGGTATCCCGACAAGGTCATGATTTCCGTTAGTCCCGTGGTCCCATGAGT | 666 |
| | | | |
| 715 | QY | gttatgaaccagatcatcttgcgaacggcggaacacc | 750 |
| | | | |
| 667 | Db | GTGTGCTCACCAAGCTTGTGCTATGTGGTGACAAC | 702 |
| | | | |

RESULT 15

| LOCUS | DEFINITION | 1692 bp | linear | PAT 29-SEP-1997 |
|-----------|--|---------|--------|-----------------|
| E02678 | E02678 cDNA encoding lipase. | | | |
| ACCESSION | E02678 | | | |
| VERSION | E02678.1 | | | |
| KEYWORDS | JP 1990299588-A/1. | | | |
| SOURCE | Galactomyces geotrichum. | | | |
| ORGANISM | Galactomyces geotrichum. | | | |
| | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |

| | |
|-----------|--|
| REFERENCE | 1 (bases 1 to 1692) |
| AUTHORS | Shimada, Y., Tomimaga, Y., Sugihara, A. and Iizumi, T. . |
| TITLE | GENE OF PROTEIN HAVING LIPASE ACTIVITY |
| JOURNAL | Patent: JP 1990299588-A 11-DEC-1990; KURITA WATER IND LTD, OSAKA CITY |
| COMMENT | OS Geotrichum candidum PN JP 1990299588-A/1 PD 11-DEC-1990 PF 27-MAR-1989 JP 1989074721 PI SHIMADA YUJI, TOMINAGA YOSHIO, SUGIHARA AKIO, IIZUMI TARO PC C12N15/55/C12N9/20, C12N15/55, C12R1:645; |

| CC | *source: clone=pgcli; | Location/Qualifiers |
|----|-----------------------|---------------------|
| FH | Key | |

| FT | sig_peptide | 1.57 | /product='s[ana] | peptide of 11base' FT |
|----|-------------|------|------------------|-----------------------|
| FT | | | | |

```

mat_peptide      58..1689
FT               /product='lipase'
FT               1..1692
FT               /product='lipase'.
FT               Location/Qualifiers
FT               1..1692
FT               /organism="Galactomyces geotrichum"
FT               /db_xref="taxon:27317"
BASE COUNT      336 a 466 c 403 g 487 t
ORIGIN

```

| | | | | | | | |
|-----------------------|--------|--------------|--------------|------------|------|--------|-------|
| Query Match | 6.2% | Score | 106.8; | DB: | 6; | Length | 1692; |
| Best Local Similarity | 56.3%; | Pred. | No. 3.1e-06; | | | | |
| Matches | 223; | Conservative | 0; | Mismatches | 167; | Indels | 6; |
| Gaps | 1; | | | | | | |

| | | | |
|----|-----|--|-----|
| QY | 361 | gagattgcctctccccaatgctgtggcccccgcgcgtgctgcgagggcgcaaatctt | 420 |
| DB | 364 | GAGGACTGCTCTACCTTTACGTTTCCGCCCCGCTGGCACCACCGCTGATGCTAAGCTC | 423 |
| QY | 421 | cccgctcctcgtctacattccgaggtgggttaogccttcggc-----gatcgagagcacc | 474 |
| DB | 424 | CCCGTCATGCTTGGATTTACGGTGGTGCCTTGTGTTTGGTCTTCTGCTCTTACGCT | 483 |
| QY | 475 | ggcagcgactttccgcgcttcaccaagcacacgggaaccaagaatggtcgtttgtaaatctc | 534 |
| DB | 484 | GGTAACGGCTACGTCACAGGAGAGTCTGGAAATGGCCAGCCTGTGTGTTTGTTCATC | 543 |
| QY | 535 | cagtcacgctctcggcagcttggtttctcgcctggccaaagccatgaaggactacgggtga | 594 |
| DB | 544 | AACCTACGTCACCGGCCCTATGGATTCTGGGTGGTATGTCATCACCAGCTAGGGTAAC | 603 |
| QY | 595 | acgaacccggcttgccttgacagcaattcgccttcgaatgggttcacacgacgcgtctcg | 654 |
| DB | 604 | ACCAACGCTGGTCTGTCACGACACCGCAAGGCTCTCAGTGGGTAGCGACCAATGGCC | 663 |
| QY | 655 | aagtcggcggaaccgcgcatcacgttacgtttggcgagttgcgagggcgaggggcc | 714 |
| DB | 664 | AACTTGGTGGTGATCCGCACACAGTCAATGATTTCCGTGGTGAAGTCCGCTGGTCCATGAGT | 723 |
| QY | 715 | gttatgaaccagatcatttgcgaacggcggaacacc | 750 |
| DB | 724 | GTTGCTCACACAGCTGTGTGCTATGGTGGTGACAACT | 759 |

Search completed: June 29, 2002, 15:03:51
Job time: 20512 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2002, 09:17:09 ; Search time 4149.72 seconds
(without alignments)
5581.286 Million cell updates/sec

Title: US-09-801-852A-1

Perfect score: 1716

Sequence: 1 atgtcttcaaccttccac.....tctcgggaacgcgggtctc 1716

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcto:*
9: gb_estl:*
10: gb_est2:*
11: gb_hct:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 80.2 | 4.7 | 620 | 9 | AW790714 D00142-R |
| 2 | 70 | 4.1 | 1800 | 9 | AA415091 MG0026 RC |
| 3 | 68 | 4.0 | 727 | 10 | BI221206 602939553 |
| 4 | 67.8 | 4.0 | 625 | 9 | AA202160 LD02542.5 |
| 5 | 67.2 | 3.9 | 810 | 10 | BF385648 602047302 |
| 6 | 66.4 | 3.9 | 875 | 10 | BF785601 602112390 |
| 7 | 65.8 | 3.8 | 783 | 12 | CNS01LTS |
| 8 | 65 | 3.8 | 526 | 10 | BI460262 603201884 |
| 9 | 64.4 | 3.8 | 725 | 10 | BG867060 602786057 |
| 10 | 64 | 3.7 | 973 | 10 | BI554561 603235901 |
| 11 | 62.6 | 3.6 | 1942 | 11 | BC015286 Mus muscu |
| 12 | 62 | 3.6 | 455 | 9 | AI508950 vc28a03.y |
| 13 | 62 | 3.6 | 564 | 10 | BE490305 |
| 14 | 62 | 3.6 | 676 | 10 | BI564550 RH61852.5 |
| 15 | 62 | 3.6 | 730 | 10 | BF505095 AT07260.5 |
| 16 | 62 | 3.6 | 1101 | 12 | CNS0027J |
| 17 | 61.4 | 3.6 | 468 | 9 | AW512144 xu54e11.x |

| | | | | | |
|----|------|-----|------|----|--------------------|
| 18 | 60.4 | 3.5 | 626 | 9 | AI527908 |
| 19 | 59.6 | 3.5 | 673 | 10 | BI574042 |
| 20 | 59.2 | 3.4 | 649 | 9 | AA246278 |
| 21 | 58.8 | 3.4 | 538 | 10 | BF256789 HVSMEF001 |
| 22 | 57.6 | 3.4 | 589 | 10 | BG263795 |
| 23 | 57.6 | 3.4 | 767 | 10 | BG970699 |
| 24 | 57.6 | 3.4 | 791 | 10 | BG969645 |
| 25 | 57.2 | 3.3 | 568 | 9 | AI516721 |
| 26 | 56.8 | 3.3 | 958 | 12 | CNS001Q9 |
| 27 | 56.6 | 3.3 | 805 | 10 | BI332322 |
| 28 | 56.4 | 3.3 | 566 | 10 | BI626691 |
| 29 | 56.4 | 3.3 | 959 | 3 | BE636588 |
| 30 | 55.6 | 3.2 | 700 | 9 | AI527918 |
| 31 | 55.4 | 3.2 | 383 | 9 | AA230461 |
| 32 | 55.2 | 3.2 | 722 | 9 | AU003170 |
| 33 | 55.2 | 3.2 | 1770 | 3 | BE636721 |
| 34 | 54.8 | 3.2 | 242 | 9 | AA377049 |
| 35 | 54.8 | 3.2 | 477 | 10 | BM374730 |
| 36 | 54.6 | 3.2 | 580 | 10 | BE402096 |
| 37 | 54.4 | 3.2 | 516 | 9 | AI108156 |
| 38 | 54.4 | 3.2 | 526 | 9 | AI108080 |
| 39 | 54.4 | 3.2 | 612 | 10 | BI331296 |
| 40 | 54.4 | 3.2 | 702 | 10 | BI146926 |
| 41 | 54.4 | 3.2 | 716 | 10 | BI220765 |
| 42 | 54.4 | 3.2 | 738 | 10 | BI332320 |
| 43 | 54.4 | 3.2 | 771 | 10 | BI144719 |
| 44 | 54.4 | 3.2 | 778 | 9 | AI062034 |
| 45 | 54.4 | 3.2 | 778 | 10 | BI328564 |

ALIGNMENTS

RESULT 1

AW790714 620 bp mRNA linear EST 01-MAY-2001
LOCUS D00142-R Lambda zap, Stratagene Blumeria graminis f. sp. hordei
DEFINITION cDNA clone D00142 similar to lipase 2 precursor, mRNA sequence.
ACCESSION AW790714 GI:13902311
VERSION AW790714.1
KEYWORDS EST.
SOURCE Blumeria graminis f. sp. hordei.
ORGANISM Blumeria graminis f. sp. hordei.
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
Erysiphales; Erysiphaceae; Blumeria.
AUTHORS 1 (bases 1 to 620)
Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and Oliver,R.P.

Gene identification in the fungal pathogen Blumeria graminis by expressed sequence tag analysis
Unpublished (2000)
Contact: Rasmussen,S.W.
Department of Yeast Genetics
Carlsberg Laboratory
10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark
Tel: 45 3327 5230
Fax: 45 3327 4766
Email: sw@ercr.dk

POLYA=No. Location/Qualifiers
1..620
/organism="Blumeria graminis f. sp. hordei"
/db_xref="taxon:62688"
/clone="D00142"
/clone_lib="Lambda zap, Stratagene"
/cell_type="Conidia"
/lab_host="Hordeum vulgare"
BASE COUNT 148 a 165 c 147 g 160 t
ORIGIN

Query Match 4.7%; Score 80.2; DB 9; Length 620;

| | | | | | | | | | |
|---------|------|--------------|----|------------|------|--------|-----|------|----|
| Matches | 192; | Conservative | 0; | Mismatches | 148; | Indels | 16; | Gaps | 2; |
|---------|------|--------------|----|------------|------|--------|-----|------|----|

| | | | |
|-----|----|---|-----|
| 351 | QY | acaggcgctcgaagattgcctctctctcaatgtcgtttgccccccgcgcctcgtcgcgagg | 410 |
| | | | |
| 339 | Db | ACAGTTTCTGAAGACATGCCCTACCTACCGAATATTTACACTCCTGCTGACTTTGACAAAGAA | 398 |
| | | | |
| 411 | QY | cgacaattctcccgctcctcgtctacattcaacggaggtgggtacgccttcggcgcgatcgag | 470 |
| | | | |
| 399 | Db | CAGCAGACTACCAAGTGATGTTGGATCCATGGAGGTGGACTGCTGGTGGCGCAGGACATC | 458 |
| | | | |
| 471 | QY | caocggcagcgaattgcgcgccttcacaacgacacagcgggaaccaagatggtctgtgtaaa | 530 |
| | | | |
| 459 | Db | CACCTATGATGGACTGGCCCTCTGTGCCCATGAAAAATG-----TGGTGGGTGGTGAC | 509 |
| | | | |
| 531 | QY | tctccagtaocgctctcgcgcagctttggtttctcgcgtggccaagccatgtagggactacgg | 590 |
| | | | |
| 510 | Db | CATTCACTATCGCCTTGGCATCTGGGATTCCTT-----CAGCACTGGGATGAACACAG | 562 |
| | | | |
| 591 | QY | tgtaaagcaacgcgcgcttgctgaacagcaattcgccttcaatgggtttcaacagcaagt | 650 |
| | | | |
| 563 | Db | TCGGGGAAATCGGGTCACTTGGACCAAGTGGCTGCTACTAGCTGGGTCCAGGACAAAT | 622 |
| | | | |
| 651 | QY | ctcgaagttcggcgcgaacccccgatacgtttacagatttggggcgcgagtctgcagcgg | 706 |
| | | | |
| 623 | Db | TGCCAACTTTGGGGGCAACCCAGCTTCGGGTGACCATCTTTGGAGAGTCAGCAGAGAG | 678 |
| | | | |

RESULT 6

| | | | | | | | |
|------------|---|--------------------|--------|----------|----------------|--------|-----------------|
| BF785601 | BF785601 | 5', mRNA sequence. | 875 bp | musculus | NCI_CGAP_K1d14 | linear | EST 12-JAN-2001 |
| LOCUS | 602112330F1 | | | | | | IMAGE:4240463 |
| DEFINITION | | | | | | | |
| ACCESSION | BF785601 | | | | | | |
| VERSION | BF785601.1 | GI:12090637 | | | | | |
| KEYWORDS | EST. | | | | | | |
| SOURCE | house mouse. | | | | | | |
| ORGANISM | Mus musculus | | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | | |
| AUTHORS | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | | | | |
| TITLE | NIH-MGC http://mgc.nci.nih.gov/. | | | | | | |
| JOURNAL | National Institutes of Health, Mammalian Gene Collection (MGC) | | | | | | |
| COMMENT | Unpublished (1999) | | | | | | |
| | Contact: Robert Strausberg, Ph.D. | | | | | | |
| | Email: cgapbs-i@mail.nih.gov | | | | | | |
| | Tissue Procurement: Jeffrey E. Green, M.D. | | | | | | |
| | cDNA Library Preparation: Life Technologies, Inc. | | | | | | |
| | cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) | | | | | | |
| | DNA sequencing by: Incyte Genomics, Inc. | | | | | | |
| | Clone distribution: MGC clone distribution information can be | | | | | | |
| | found through the I.M.A.G.E. Consortium/LLNL at: | | | | | | |
| | http://image.llnl.gov | | | | | | |
| | Plate: LLAM9855 row: c column: 24 | | | | | | |
| | High quality sequence stop: 757. | | | | | | |

| | |
|-----------------|---------------|
| FEATURES | SOURCE |
|-----------------|---------------|

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4240463"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"
198 a 239 c 239 q 209 t
BASE COUNT

```

RESULT 7

| | | | | | |
|-----------|--|------------|-----|--------|-----------------|
| CNSOILTS | CNSOILTS | 783 bp | DNA | linear | GSS 14-JUN-2001 |
| LOCUS | Anopheles gambiae GSS SP6 end of clone 20C01 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence. | | | | |
| ACCESSION | AL150081 | | | | |
| VERSION | ALI50081.1 | GI:7010560 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | African malaria mosquito. | | | | |
| ORGANISM | Anopheles gambiae | | | | |
| | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Anopterygota; Diptera; Nematocera; Culicoidae; Anopheles. | | | | |
| REFERENCE | 1 (bases 1 to 783) | | | | |
| AUTHORS | Genoscope. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr) | | | | |
| | - Web : www.genoscope.cns.fr | | | | |
| REFERENCE | 2 (bases 1 to 783) | | | | |
| AUTHORS | Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France | | | | |
| COMMENT | This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur. | | | | |

FEATURES
SOURCE

```

source
i. . /03
/organism="Anopheles gambiae"
/strain="PESt"
/db_xref="taxon:7165"
/clone="20C01"
/clone_lib="NotreDamel"
/notes"end : SP6"
169 a 214 C 179 t 3 others
BASE COUNT.
ORIGIN

```

Query Match
Best Local

Qy 361 gagagattgcctctcctcaatgcttggttccccccggcgctgtagcgaggggacaaatttt 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Best Local Similarity 36.89; P-Id: NO. 0.17; Indels 9; Gaps 2;
Matches 195; Conservative 0; Mismatches 167;

```

Db 137 GAGAAATTGCTCTACTTGAATGTGTACCGCCGCCCCAAAGGCAACGGCAGTGCAGTCTCCCTG 196
QY 421 ccgcctcctctacattcagaggtgctagccttcgagcgtgagcaccgagc 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 CCGTAATGTTGTATACATAGTGTGGTGGATATTTCTTCGGTCCGGCGATCCACACTG 256
QY 481 gactttgcgccttcacccaagcacacgggaaccaagatggtcgttgaatatctccagtac 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 TACGGGCGGAGCGAATCTTGGCCAC--AAAACAAGTGATCCCTCGTCACACTCCAGTAT 313
QY 541 cgtctcggagcgttggtttcttcctcgtggccaagccatgaaggactcaggtgtaacaagac 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 314 CGTCTCGGTGTGTAGCTTCTCTTCACGGGGGATGCACACGCTACCGGT-----AAC 367
QY 601 gccggtctgttaccagcaattcccttcaatgggttcaaacagcagctctcgaagtac 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 368 TACGGTATGCTTGATCAAGTGTAGCCCTCGATGGTGGTCAATCAGACATCGGCGCCTTC 427
QY 661 ggccgcaaccccgatcacgttaccgattggggcgagctctcagcgccaggtccgtttatg 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 428 GGTGTGATCTCACTCGGTGACACTGTTCGGTGAAGTGGCGCGGTGCTTCGTGTCAG 487
QY 721 aaccagatcat 731
    |||||
Db 488 CTGCACATGAT 498

RESULT 8
BI460262 BI460262 526 bp mRNA linear EST 21-AUG-2001
LOCUS 603201884F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267757 5',
DEFINITION mRNA sequence.
ACCESSION BI460262
VERSION BI460262.1 GI:15250918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 526)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11675 row: g column: 22
High quality sequence stop: 525.
FEATURES
source
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5267757"
/lab_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
77 a 163 c 187 g 99 t

```

```

Query Match 3.8%; Score 65; DB 10; Length 526;
Best Local Similarity 52.8%; Pred. No. 0.22;
Matches 196; Conservative 0; Mismatches 160; Indels 15; Gaps 2;

QY 361 gaggattgcctctcctcaatgctgttgcctcccgccgctgctgctgagggcgacaaatctt 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123 GAGGACTGTCTGTACCTGAACGTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 182
QY 421 ccgcctcctctacattcagaggtggtacagccttcgagcgtgagcaccgagc 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183 CCAGTGTATGTTCTGGTTCCTCGGGAGGCGCCTTCATCTGGTGGCGCTGCTTCTTCTGACAG 242
QY 481 gactttgcgccttcacccaagcacacgggaaccaagatggtcgttgaatatctccagtac 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 GGCTCTGACTGTGGCGCGCGCGGAGAAAG-----TGGTGTGTGGTGTCTTCGACGAC 293
QY 541 cgtctcggagcgttggtttcttcctcgtgccaagccatgaaggactcaggtgtaacaagac 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 AGGCTCGGCATCTTCGGCTTCCTGAGCACGACGACACGACGACGACGACGACGACGACG 347
QY 601 gccggtctgttaccagcaattcccttcaatgggttcaaacagcagctctcgaagtac 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 TGGGGGCTGCTGGACCATGGCGGCTCTGCTGGTGGTGGAGAAACATCGGACGCTTC 407
QY 661 ggccgcaaccccgatcacgttaccgattggggcgagctctcagcgccaggtccgtttatg 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 GGGGAGACCCAGCAATGTGACCTCTTCGGCCAGTCGGCGGGGCCATGAGCATCTCA 467
QY 721 aaccagatcat 731
    |||||
Db 468 GGACTGATGAT 478

RESULT 9
BG867060 BG867060 725 bp mRNA linear EST 29-MAY-2001
LOCUS 602786057F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4912023 5',
DEFINITION mRNA sequence.
ACCESSION BG867060
VERSION BG867060.1 GI:14217600
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10815 row: a column: 16
High quality sequence stop: 724.
FEATURES
source
1..725
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4912023"
/lab_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: salivary gland; Vector: pcMW-SPO6T6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life

```



```

FEATURES
source
location/Qualifiers
1. .973
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5309856"
/clone_lib="NCI_CGAP_L19"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.9 Kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."

```

```

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 25 Row: 1 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein
This clone has the following problem: frame shifted.
Location/Qualifiers

```



```

Db 662 TTCAATGCGATCCC 676
|||||
RESULT 15
BF505095 730 bp mRNA linear EST 19-APR-2001
LOCUS AT07260.5prime AT: Drosophila melanogaster adult testes potB7
DEFINITION Drosophila melanogaster cDNA clone AT07260 5 similar to alpha-Est6:
Fban001108 'enzyme' located on: 3R 84D5-84D5;; 04/07/2001, mRNA
sequence.
ACCESSION BF505095
VERSION BF505095.2 GI:13687701
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera,
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 730)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
Liao,G., Miranda,A., Mista,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,
Celniker,S. and Rubin,G.M.
BDGP/HMI AT Drosophila EST Project
JOURNAL Unpublished (2000)
COMMENT On Dec 6, 2000 this sequence version replaced gi:11568396.
Contact: Stapleton, M.
BDGP

```

Search completed: June 29, 2002, 13:05:40
Job time: 13711 sec

```

FEATURES
source
1..730
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT07260"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120; DH5-alpha. Plates
AT.121-AT.319; DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: potB7; Site_1: EcoRI;
Site_2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into potB7. Plasmid cDNA library."
BASE COUNT 166 a 179 c 207 g 178 t
ORIGIN
Query Match 3.6%; Score 62; DB 10; Length 730;
Best Local Similarity 53.0%; Pred. No. 0.71;
Matches 159; Conservative 0; Mismatches 135; Indels 6; Gaps 1;
QY 417 tttccctccctcattcaccaggggtggtcgcgttcggtcgcgagcagcagcgg 476
|||||
Db 358 TCTGCCCGTAATGGTGATATCTACGGGGTGCCTTCACAGCGGGAGCTGTGAAGAG 417
|||||
QY 477 cagcgactttgcgccttcaccagcacacaggggaaccaagatggtcgttgtaaatctoca 536
|||||
Db 418 CAATACGGCGCGGATTACTTGATGAGCAGGG-----ATGTTGTATATGCTATTCAA 471
|||||
QY 537 gtacgcgtctcgacgtttgttctcgtcgtgcacagccatgaaggactacgggtgaac 596
|||||
Db 472 CTATCGCCTCTCTCCCTGGGATCTCTTAGTATGCCAGAGTGCAAGTTGGATGTCGGGG 531
|||||

```

```

QY 597 gaacgcggcttggcttgaccaccgaattcgcccttcaatgggttcaacagcagcgtctcgaa 656
|||||
Db 532 CAATGCGGGTCTCCATGATCAGTTACTAGCCCTTCAATGGGTAGCCAGCACATACGGA 591
|||||
QY 657 gttcgggggcaaccccgatcacgtttacgatttggggcgagttcgcagcgcagggtcggt 716
|||||
Db 592 CTTCAACGGAGATCCGCGAATATATACGCTTTTGGCGAAAGTGGGGGGCGGCTCGGT 651
|||||

```

